

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
26 July 2001 (26.07.2001)

PCT

(10) International Publication Number
WO 01/53531 A3

(51) International Patent Classification⁷: **C12Q 1/68**, C12N 15/12, C07K 14/47, C12N 5/10, A01K 67/027 // A61P 19/12

(74) Agents: **FORBES, James, C. et al.**; Pharmacia Corporation, Corporate Patent Department, P.O. Box 5110, Chicago, IL 60680 (US).

(21) International Application Number: PCT/US01/00016

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(22) International Filing Date: 18 January 2001 (18.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/176,523 18 January 2000 (18.01.2000) US

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (for all designated States except US): **PHARMACIA CORPORATION [US/US]**; Corporate Patent Department, P.O. Box 5110, Chicago, IL 60680 (US).

Published:

— with international search report

(88) Date of publication of the international search report:
20 December 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PHIPPARD, Deborah** [GB/US]; 16938 Lewis Spring Farms, Wildwood, MO 63005 (US). **VASANTHAKAMUR, Geetha** [US/US]; 1176 Claytonia Terrace, St. Louis, MO 63117 (US). **DOTSON, Stanton** [US/US]; 379 Palomino Hill Court, Chesterfield, MO 63005 (US). **MA, Xiao-Jun** [CN/US]; 4482 Calle Mar De Armonia, San Diego, CA 92130 (US).

WO 01/53531 A3

(54) Title: OSTEOARTHRITIS TISSUE DERIVED NUCLEIC ACIDS, POLYPEPTIDES, VECTORS, AND CELLS

(57) Abstract: This invention relates to nucleic acids derived from osteoarthritis (OA) tissue as well as compounds and compositions made using these nucleic acids. A variety of nucleic acid and polypeptide compounds and compositions are described and specifically disclosed. The nucleic acids or polypeptides may be contained within vectors or host cells and then used to produce agents such as nucleic acids, polypeptides, fragments of polypeptides, antibodies, and variants of each. These molecules can be used to diagnose or treat osteoarthritis, or to analyze the disease-modifying activity of OA drugs. Cells containing one or more nucleic acids or polypeptides of the invention can also be used as targets in high-throughput screening methods, particularly in screening for compounds designed to identify compositions affecting osteoarthritis.

BEST AVAILABLE COPY

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/00016

A. CLASSIFICATION OF SUBJECT MATTER				
IPC 7 C12Q1/68 C12N15/12 C07K14/47 C12N5/10 A01K67/027 //A61P19/12				
According to International Patent Classification (IPC) or to both national classification and IPC				
B. FIELDS SEARCHED				
Minimum documentation searched (classification system followed by classification symbols)				
IPC 7 C12Q C12N A01K C07K				
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched				
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)				
BIOSIS				
C. DOCUMENTS CONSIDERED TO BE RELEVANT				
Category	Citation of document, with indication, where appropriate, of the relevant passages			Relevant to claim No.
X	WO 99 31233 A (SCHOR SETH LAWRENCE ;UNIV DUNDEE (GB); SCHOR ANA MARIA (GB)) 24 June 1999 (1999-06-24) figure 1 page 6, line 9 -page 18, last line page 23, line 21 -page 25, line 5 page 34, line 33 -page 35, line 16 claims; examples --- LUST G ET AL: "FIBRONECTIN AS A MARKER FOR OSTEOARTHRITIS" JOURNAL OF RHEUMATOLOGY, vol. 14, no. SUPPL. 14, 1987, pages 28-29, XP001002521 ISSN: 0315-162X the whole document --- -/--			2-15,17, 18,21-31 6,30 6,30 -/-
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.				
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority, claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed				
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family				
Date of the actual completion of the international search		Date of mailing of the international search report		
12 June 2001		12.09.01		
Name and mailing address of the ISA		Authorized officer		
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		ANDRES S.M.		

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 01/00016

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EM_EST [Online] EMBL; Accession number : AA375384 (ID : HSZ80510), 18 April 1997 (1997-04-18) "EST87654 HSC172 cells II Homo sapiens cDNA 5'" XP002168641 the whole document & ADAMS, M. ET AL.: "Initial assessment of human gene diversity and expression patterns based upon 83 million basepairs of cDNA sequence." NATURE., vol. 377, 1995, pages 3-174, XP002920293 ISSN: 0028-0836</p> <p>---</p>	2-4, 10-12, 17, 21-25,31
A	<p>WRIGHT G D ET AL: "Association of two loci on chromosome 2q with nodal osteoarthritis." ANNALS OF THE RHEUMATIC DISEASES, vol. 55, no. 5, 1996, pages 317-319, XP001002233 ISSN: 0003-4967 the whole document</p> <p>---</p>	1-6
A	<p>AIGNER THOMAS ET AL: "Reexpression of type IIA procollagen by adult articular chondrocytes in osteoarthritic cartilage." ARTHRITIS & RHEUMATISM, vol. 42, no. 7, July 1999 (1999-07), pages 1443-1450, XP001002223 ISSN: 0004-3591</p> <p>---</p>	
A	<p>MYERS STEPHEN L: "Synovial fluid markers in osteoarthritis." RHEUMATIC DISEASE CLINICS OF NORTH AMERICA, vol. 25, no. 2, May 1999 (1999-05), pages 433-449, XP001002541 ISSN: 0889-857X</p> <p>---</p>	
P,X	<p>WO 00 49159 A (TERUMO CORP) 24 August 2000 (2000-08-24)</p> <p>page 14, line 23 -page 19, line 12 example 1</p> <p>-----</p>	2-4, 10-15, 23-27,31

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/00016

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
Although claim 30 (as far as an in vivo method is concerned) is directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. Claims Nos.: 20 because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Invention 1 : Claims 1-31 (all partially)

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claim : Invention 1. : Claims 1 to 31 (all partially)

A nucleic acid consisting of SEQ ID 1; derivatives thereof; vectors, cells and animals expressing it; a computer-readable media recording it; its uses in the diagnosis or prognosis of osteoarthritis.

2. Claim : Inventions 2. to 82. : Claims 1 to 31 (all partially)

As for subject 1, but concerning SEQ IDs 2 to 82 respectively.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 20

* Present claim 20 relates to a compound defined by reference to a desirable characteristic or property, namely being capable of interacting with a protein according to claim 13.

The claims cover all compounds having this characteristic or property, whereas the application does not provide support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for any of such compounds. In the present case, the claim so lack support, and the application so lacks disclosure, that a meaningful search is impossible. Independent of the above reasoning, the claim also lack clarity (Article 6 PCT). An attempt is made to define the compound by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search impossible. Consequently, no search has been carried out for claim 20.

* Claim 27 is drawn to a labeled antibody with reference to an unknown claim. As claim 15 is the only other claim concerning an antibody, claim 27 has been read and searched as referring to claim 15.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 01/00016

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9931233	A 24-06-1999	AU 1571299	A 05-07-1999	EP 1042466 A 11-10-2000
WO 0049159	A 24-08-2000	AU 2575000	A 04-09-2000	-----



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
26 July 2001 (26.07.2001)

PCT

(10) International Publication Number
WO 01/53531 A2

(51) International Patent Classification⁷: **C12Q 1/68**, C12N 15/12, C07K 14/47, C12N 5/10, A01K 67/027 // A61P 19/12

(74) Agents: **FORBES, James, C. et al.**; Pharmacia Corporation, Corporate Patent Department, P.O. Box 5110, Chicago, IL 60680 (US).

(21) International Application Number: **PCT/US01/00016**

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(22) International Filing Date: 18 January 2001 (18.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/176,523 18 January 2000 (18.01.2000) US

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PHIPPARD, Deborah** [GB/US]; 16938 Lewis Spring Farms, Wildwood, MO 63005 (US). **VASANTHAKAMUR, Geetha** [US/US]; 1176 Claytonia Terrace, St. Louis, MO 63117 (US). **DOTSON, Stanton** [US/US]; 379 Palomino Hill Court, Chesterfield, MO 63005 (US). **MA, Xiao-Jun** [CN/US]; 4482 Calle Mar De Armonia, San Diego, CA 92130 (US).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/53531 A2

(54) Title: OSTEOARTHRITIS TISSUE DERIVED NUCLEIC ACIDS, POLYPEPTIDES, VECTORS, AND CELLS

(57) **Abstract:** This invention relates to nucleic acids derived from osteoarthritis (OA) tissue as well as compounds and compositions made using these nucleic acids. A variety of nucleic acid and polypeptide compounds and compositions are described and specifically disclosed. The nucleic acids or polypeptides may be contained within vectors or host cells and then used to produce agents such as nucleic acids, polypeptides, fragments of polypeptides, antibodies, and variants of each. These molecules can be used to diagnose or treat osteoarthritis, or to analyze the disease-modifying activity of OA drugs. Cells containing one or more nucleic acids or polypeptides of the invention can also be used as targets in high-throughput screening methods, particularly in screening for compounds designed to identify compositions affecting osteoarthritis.

OSTEOARTHRITIS TISSUE DERIVED NUCLEIC ACIDS, POLYPEPTIDES, VECTORS, AND CELLS

Field of the Invention

5 This invention relates to nucleic acids derived from osteoarthritis (OA) tissue as well as compounds and compositions made using these nucleic acids. A variety of nucleic acid and polypeptide compounds and compositions are described and specifically disclosed. The nucleic acids or polypeptides may be contained within vectors or host cells and then used to produce agents such as
10 nucleic acids, polypeptides, fragments of polypeptides, antibodies, and variants of each. These molecules can be used to diagnose or treat osteoarthritis, or to analyze the disease-modifying activity of OA drugs. Cells containing one or more nucleic acids or polypeptides of the invention can also be used as targets in high-throughput screening methods, particularly in screening for compounds designed
15 to identify compositions affecting osteoarthritis.

Background of the Invention

Osteoarthritis (OA) is a slowly progressing degenerative disease characterized by cartilage destruction that affects one or more joints. It is the most prevalent articular disease, and can severely impair mobility and lower extremity function (Ling and Bathon, *J Am Geriatr Soc* 46:216-25 (1998)). The term osteoarthritis implies an inflammatory disease. Although inflammatory cells may be present, OA is considered to be an intrinsic disease of cartilage in which biochemical and metabolic alterations result in its breakdown. Diagnosis is typically based upon radiological examination as well as clinical observations such as localized tenderness and bony or soft tissue swelling, joint function, severity of pain and the ease with which everyday functions such as climbing stairs can be performed. Characteristic radiographic findings include subchondral bone sclerosis, subchondral cysts and osteophytosis. Although joint space narrowing is considered to be a marker for articular cartilage thinning, in patients

with early OA, who do not have radiographic evidence of bony changes, joint space narrowing alone does not accurately indicate the status of the articular cartilage. Similarly osteophytosis alone, in the absence of other radiographic features of OA, may be due to aging rather than OA. The correlation between the 5 pathological severity of OA and symptoms is poor. Many individuals with radiographic changes of advanced OA have no symptoms. The risk factors for pain and disability in subjects with OA is still poorly understood.

Estimating OA disease progression presents an even greater challenge than diagnosis. The current "gold standard" measure of disease progression is the 10 change in joint space caused by articular cartilage loss observed using plain X-rays (Mazzuca et al., *Osteoarthritis and Cartilage* 5:217-226 (1997)). Since changes are small (1-2 mm per year), a minimum of one year is required before sufficient changes have occurred to be detectable. An additional problem with this technique is the requirement that the joint must be in exactly the same 15 position for each radiological examination. Thus, there is a need in the art for improved OA diagnostic and disease progression monitoring-methods.

Current osteoarthritis therapies treat the associated pain and inflammation, but do little, if anything, to modify the course of the disease. Drug trials of potential disease modifying osteoarthritis drugs (DMOADs) require a minimum 20 of one year before drug efficacy can be assessed. Therefore there is a great need for biological markers and design assay systems that will predict disease modifying efficacy of Osteoarthritis drugs (DMOADS) within a significantly shorter time frame than the current year required for assessment of radiological changes. Also, disease markers that will detect early disease and allow early 25 intervention with pharmacologic agents are much needed, but they are presently unavailable (Ling and Bathon, *J Am Geriatr Soc* 46:216-25 (1998)).

Summary of the Invention

It is an object of the present invention to provide biological markers and diagnostic assays, which satisfy the unmet diagnostic and monitoring needs in the art. Markers of OA disease and for novel disease modifying therapeutics are 5 described in this invention. These nucleic acid sequences have been identified using detailed genomics and bioinformatics analyses combined with our understanding of the pathology of OA. This invention identifies genes involved in the biology of OA and representing targets from which diagnostic tools, such as monoclonal antibodies, can be generated to aid in monitoring treatment, 10 definition and diagnosis of OA as well as new targets for therapeutic drug intervention.

Exemplary Uses of the Invention

The present invention provides nucleic acids and polypeptide sequences that can be used to generate diagnostic tools and to develop small molecule drugs. 15 The sequences of the present invention can also be used for the direct generation of therapeutics. Nucleic acid sequences of the present invention also represent probes and primers which facilitate cloning of genes.

Useful for generating diagnostic reagents

Nucleic acid probes, proteins, and polypeptides made using the nucleic 20 acids of the present invention or derived from the sequence information of these nucleic acids are disclosed.

a) Sequences represent markers that correlate with OA disease. Where ESTs are identified as coding for previously identified genes, full 25 length sequence can be cloned (appropriate purification tags added if desired), expressed in a bacterial, yeast, insect, mammalian or other cell, and the recombinant protein used to raise antibodies which will have utility in OA diagnosis and may aid in analyzing DMOAD efficacy.

- b) Sequences representing known genes that are now shown to correlate with OA disease allow commercially produced antibodies to these genes, raised for unrelated purposes, to be evaluated as markers of osteoarthritis disease and their response to drug treatment studied.
- c) The full length sequences of the ESTs identified can also be directly injected into animals to elicit an antibody response.
- d) Antibodies generated by any of the methods described above can be used to generate ELISAs (Enzyme linked immunosorbent assays) for the detection of OA serum or synovial fluid markers (or other appropriate body tissues and fluids e.g. saliva) and their response to drug treatment. The same antibodies are also useful for detecting OA and DMOAD markers using Western blotting techniques and other antibody-based diagnostics.
- e) As described in a) proteins can be generated to the genes represented by the identified ESTs. This process will also generate positive controls to test the sensitivity and efficacy of the ELISAs described in d).
- f) Where full length sequences are known or obtained, biochemical assays for the protein product can be designed to generate an assay for OA.

g) The sequences specify oligonucleotides which can be used in polymerase chain reaction (PCR) detection procedures to determine the level of gene expression which infers the cellular state or gene response.

5

h) PCR primer and probe sets homologous to the identified ESTs can be designed for use in quantitative PCR technology such as the TaqMan technology used in the ABI7700, therefore genes identified here as regulated as a result of OA can be quantified in human patients.

10

i) Sequences can be used to clone their orthologous genes (if present) in species other than Homo sapiens. This allows their expression to be studied in animal models of OA such as the anterior cruciate ligament dog model, the rabbit partial meniscectomy model and the guinea pig spontaneous model.

15

j) The DNA clones of these sequences can be spotted onto an array or microarray by themselves or as a set or in combination with other genes to determine the level of gene expression which can be used to infer a cellular state or cell response.

20

k) Sequences can be placed on a computer-readable medium, and can be used to identify another nucleic acid by comparing nucleotide sequence information using computerized means.

25

Useful as targets for small molecule drug development

The present invention also provides methods and assays to discover compounds and compositions related to OA.

5 a) Sequences represent genes which can be cloned, over-expressed in a bacterial, yeast, insect, mammalian or other cell, and the active protein used in high throughput screening for novel inhibitors. For example, a biologically-active compound or composition may be identified by examining the interaction between a protein, polypeptide, or fragment of the invention and a compound or composition, and comparing this to a similar interaction of the protein, polypeptide, or fragment of the invention with a control.

10 b) Compounds and compositions which are detectable through screens using sequences that represent genes.

15 c) Sequences can be used to clone their promoter regions, which in turn can be linked to a reporter gene, such as luciferase, and the resultant recombinant reporter constructs used to screen compounds which alter gene expression.

20 d) The sequences can be used to identify transcription factors which modulate their expression. These transcription factors can be cloned, over-expressed in a bacterial, yeast, insect, mammalian or other cell, and the active transcription factor used in high throughput screening for small molecule inhibitors of gene expression.

25 e) Sequences represent cellular markers correlated to a cellular state which individually or in combination can be measured in response to a compound for those compounds which either suppress or

activate the genes and thus alter the state of the cell in a desired manner.

Useful for the direct generation of therapeutics

- a) The sequences represent genes which can be cloned, over-expressed in a bacterial, yeast, insect, mammalian or other cell, and the encoded protein used as a protein therapeutic.
- 5 b) The sequences represent genes which can be directly injected to elicit antibodies useful as therapeutics.
- 10 c) The sequences represent genes which can be cloned, expressed in a bacterial, yeast, insect, mammalian or other cell, and the protein used to generate antibodies useful as therapeutics.
- 15 d) The sequences can be introduced into an animal to create transgenic animals
- e) The sequences can be used to generate antisense DNA molecules useful to suppress gene expression and provide a therapeutic benefit.
- 20 f) The sequences can be used to generate antisense oligonucleotides useful to suppress gene expression and provide a therapeutic benefit.
- 25 g) The sequences can be used to generate sense DNA or sense oligonucleotides which will act by co-suppression to provide a therapeutic benefit.

h) The sequences or genes represented by the sequences are useful as gene therapy for activating or suppressing themselves, other genes or entire pathways of genes.

5

The sequences will facilitate cloning the complete gene

The present invention also includes the entire coding region, the promoter which controls gene transcription, the untranslated region which may control RNA stability and translation as well as the genomic clone containing exon and intron information.

10. a) The sequences specify oligonucleotide templates that can be used to amplify the full length gene using PCR.

15 b) The DNA clones can be labeled in a manner that they can be used to hybridize to a corresponding full length gene in order to detect and clone the full length gene.

20 c) The sequences have utility in other procedures, not limited to the two previous examples, to clone the full length gene.

Detailed Description of the Invention

A. General Concepts and Definitions

These detailed descriptions are presented for illustrative purposes only
25 and are not intended as a restriction on the scope of the invention. Rather, they are merely some of the embodiments that one skilled in the art would

understand from the entire contents of this disclosure. All parts are by weight and temperatures are in Degrees centigrade unless otherwise indicated.

Abbreviations and Definitions

The following is a list of abbreviations and the corresponding meanings as

5 used interchangeably herein:

IMDM = Iscove's modified Dulbecco's media

mg = milligram

ml or mL = milliliter

μ g or ug= microgram

10 μ l or ul = microliter

ODNs= oligonucleotides

PCR= polymerase chain reaction

RP-HPLC = reverse phase high performance liquid chromatography

15 The following is a list definitions of various terms used herein:

The term "altered" means that expression differs from the expression response of cells or tissues not exhibiting the phenotype.

The term "amino acid(s)" means all naturally occurring L-amino acids.

20 The term "biologically active" means activity with respect to either a structural or a catalytic attribute, which includes the capacity of a nucleic acid to hybridize to another nucleic acid molecule, or the ability of a protein to be bound by an antibody (or to compete with another molecule for such binding), among others. Catalytic attributes involve the capacity of the agent to mediate a chemical reaction or response.

25 The term "chromosome walking" means a process of extending a genetic map by successive hybridization steps.

The term "**cluster**" means that BLAST scores from pairwise sequence comparisons of the member clones are similar enough to be considered identical with experimental error.

5 The term "**complement**" means that one nucleic acid exhibits complete complementarity with another nucleic acid.

The term "**complementarity**" means that two molecules can hybridize to one another with sufficient stability to permit them to remain annealed to one another under conventional high stringency conditions.

10 The term "**complete complementarity**" means that every nucleotide of one molecule is complementary to a nucleotide of another molecule.

The term "**degenerate**" means that two nucleic acid molecules encode for the same amino acid sequences but comprise different nucleotide sequences.

15 The term "**exogenous genetic material**" means any genetic material, whether naturally occurring or otherwise, from any source that is capable of being inserted into any organism.

The term "**expansion**" means the differentiation and proliferation of cells.

The term "**expressed sequence tags (ESTs)**" means randomly sequenced members of a cDNA or complementary DNA library.

20 The term "**expression response**" means the mutation affecting the level or pattern of the expression encoded in part or whole by one or more nucleic acid molecules.

The term "**fragment**" means a nucleic acid molecule whose sequence is shorter than the target or identified nucleic acid molecule and having the identical, the substantial complement, or the substantial homologue of at least 7 contiguous nucleotides of the target or identified nucleic acid molecule.

25 The term "**fusion molecule**" means a protein-encoding molecule or fragment that upon expression, produces a fusion protein.

The term "**fusion protein**" means a protein or fragment thereof that comprises one or more additional peptide regions not derived from that protein. Such molecules

may be derivatized to contain carbohydrate or other moieties (such as keyhole limpet hemocyanin, etc.).

The term "**hybridization probe**" means any nucleic acid capable of being labeled and forming a double-stranded structure with another nucleic acid over a region 5 large enough for the double stranded structure to be detected.

The term "**isolated**" means an agent is separated from another specific component with which it occurred. For example, the isolate material may be purified to essential homogeneity, as determined by PAGE or column chromatography, such as HPLC. An isolated nucleic acid can comprise at least about 50, 80, or 90% (on 10 a molar basis) of all macromolecular species present. Some of these methods described later lead to degrees of purification appropriate to identify single bands in electrophoresis gels. However, this degree of purification is not required.

The term "**marker nucleic acid**" means a nucleic acid molecule that is utilized to determine an attribute or feature (e.g., presence or absence, location, correlation, 15 etc.) of a molecule, cell, or tissue.

The term "**mimetic**" refers to a compound having similar functional and/or structural properties to another known compound or a particular fragment of that known compound.

The term "**minimum complementarity**" means that two molecules can hybridize 20 to one another with sufficient stability to permit them to remain annealed to one another under at least conventional low stringency conditions.

The term "**PCR probe**" means a nucleic acid capable of initiating a polymerase activity while in a double-stranded structure with another nucleic acid. For example, Krzesicki, *et al.*, *Am. J. Respir. Cell Mol. Biol.* 16:693-701 (1997), 25 incorporated by reference in its entirety, discusses the preparation of PCR probes for use in identifying nucleic acids of osteoarthritis tissue. Other methods for determining the structure of PCR probes and PCR techniques have been described.

The term "**phenotype**" means any of one or more characteristics of an organism, tissue, or cell.

The term "**polymorphism**" means a variation or difference in the sequence of the gene or its flanking regions that arises in some of the members of a species.

5 The term "**primer**" means a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (e.g., in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a
10 primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but must be sufficiently complementary to hybridize with a template.

15 The term "**probe**" means an agent that is utilized to determine an attribute or feature (e.g. presence or absence, location, correlation, etc.) of a molecule, cell, tissue, or organism.

The term "**product score**" refers to a formula which indicates the strength of a BLAST match using the fraction of overlap of two sequences and the percent
20 identitiy. The formula is as follows:

$$\text{Product Score} = \frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length}(\text{Seq1}), \text{length}(\text{Seq2})\}}$$

25 The term "**promoter region**" means a region of a nucleic acid that is capable, when located in *cis* to a nucleic acid sequence that encodes for a protein or peptide, of functioning in a way that directs expression of one or more mRNA molecules.

The term "**protein fragment**" means a peptide or polypeptide molecule whose amino acid sequence comprises a subset of the amino acid sequence of that protein.

5 The term "**protein molecule/peptide molecule**" means any molecule that comprises five or more amino acids.

The term "**recombinant**" means any agent (e.g., DNA, peptide, etc.), that is, or results from, however indirectly, human manipulation of a nucleic acid molecule.

The recombination may occur inside a cell or in a tube.

10 The term "**selectable or screenable marker genes**" means genes whose expression can be detected by a probe as a means of identifying or selecting for transformed cells.

The term "**singleton**" means a single clone.

The term "**specifically bind**" means that the binding of an antibody or peptide is not competitively inhibited by the presence of non-related molecules.

15 The term "**specifically hybridizing**" means that two nucleic acid molecules are capable of forming an anti-parallel, double-stranded nucleic acid structure.

The term "**substantial complement**" means that a nucleic acid sequence shares at least 80% sequence identity with the complement.

20 The term "**substantial fragment**" means a fragment which comprises at least 100 nucleotides.

The term "**substantial homologue**" means that a nucleic acid molecule shares at least 80% sequence identity with another.

25 The term "**substantial identity**" means that 70% to about 99% of a region or fragment in a molecule is identical to a region of a different molecule. When the individual units (e.g., nucleotides or amino acids) of the two molecules are schematically positioned to exhibit the highest number of units in the same position over a specific region, a percentage identity of the units identical over the total number of units in the region is determined. Numerous algorithmic and

computerized means for determining a percentage identity are known in the art. These means may allow for gaps in the region being considered in order to produce the highest percentage identity.

5 The term "**substantially hybridizing**" means that two nucleic acid molecules can form an anti-parallel, double-stranded nucleic acid structure under conditions (e.g. salt and temperature) that permit hybridization of sequences that exhibit 90% sequence identity or greater with each other and exhibit this identity for at least a contiguous 50 nucleotides of the nucleic acid molecules.

10 The term "**substantially purified**" means that one or more molecules that are or may be present in a naturally occurring preparation containing the target molecule will have been removed or reduced in concentration.

The term "**tissue sample**" means any sample that comprises more than one cell.

B. Data Mining and Bioinformatic Computer Analysis

The present invention relates to osteoarthritis tissue-derived nucleic acid sequences SEQ NOS. 1-82. These targets were identified by bioinformatic computer analysis and data mining of expressed sequence tags (ESTs) derived from sequencing cDNAs from normal and OA diseased tissues. The data mining effort used sequence comparison techniques (based on BLAST computer program comparisons of individual ESTs) to evaluate which ESTs were preferentially observed in the target libraries versus control and/or normal libraries. The selected set of sequences represent a group of genes encoding multiple molecular targets for OA diagnostics. This same set of sequences also represents targets for new OA therapeutic drugs.

The present invention also relates to sequences derived from SEQ NOS. 1-25 82. A variety of computerized means for identifying sequences derived from the SEQ NOS. 1-82 exists. These include the five implementations of BLAST, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN), as well

as FASTA and others (Coulson, *Trends in Biotechnology* 12:76-80 (1994); Birren *et al.*, *Genome Analysis* 1:543-559 (1997)).

The BLASTXis one such program that allows the comparison of nucleic acid sequences in the present invention to protein databases to search for the presence of protein homologs. The strength of a BLAST identity match is indicated by the "product score". This score is a normalized value between 0 and 100, with 100 indicating 100% identity over the entire length of the shorter of the two sequences, and 0 representing no shared identity between the sequences.

Other programs which use either individual sequences or make models from related sequences to further identify sequences derived from SEQ NO 1- SEQ NO 82 exist. Model building and searching programs includes HMMer (Eddy), MEME (Bailey and Elkan, *Ismb* 3: 21-29 (1995)) and PSI-BLAST (Altschul *et al.*, *Nucleic Acids Res* 25: 3389-3402 (1997)). Another set of programs which use predicted, related, or known protein structures to further identify sequences derived from SEQ NO 1- SEQ NO 82 exists. Structure-based searching programs includes ORF and PROSITE (Bairoch, A. *Nucleic Acid Res.* 20(Suppl.):2013-18 (1992)). Other programs which use individual sequences or related groups of sequences rely on pattern discovery to further identify sequences derived from SEQ NO:1- 82 exist. Pattern recognition programs include Teiresias (Rigoutsos, I. and A. Floratos, *Bioinformatics* 14:55-67 (1998)), Genotator (Harris, N. *Genome Res.* 7:754-762 (1997)).

For identification of protein structure, SPASM, RIGOR, and many other programs exist (Kleywelt, G. *J. Mol. Biol.* 285:1887-97 (1999)). Guermeur *et al.* *Bioinformatics* 15:413-21 (1999)). These programs can search any appropriate database, such as but not limited to GenBank, dbEST, EMBL, SwissProt, PIR, and GENES, or protein databases such Protein DataBank (PDB), SCOP, CATH, the Protein Information Resource (PIR) International Protein Sequence Database, the RESAID Database, and FSSP. Computerized means for designing modifications

in protein structure are also known in the art (Dahiyat and Mayo, *Science* 278:82-87 (1997)). Additional information about protein characteristics also can be predicted, such as detecting signal peptides which are secreted (SignalP program) (Nielsen et al., *Protein Engineering* 12:3-9 (1999)).

5 The following protein or polypeptide embodiments of the invention can be identified through assays known in the art, including high throughput screening assays. The proteins or polypeptides possess a detectable activity in a functional assay and can be identified by that functional assay. For example, a kinase activity assay is discussed in U.S. Patent 5,759,787, and the references therein.

10 **C. Agents of the Invention**

(a) Nucleic Acids

Agents of the present invention include nucleic acids and, more specifically, osteoarthritis tissue-derived nucleic acids. A subset of the nucleic acid molecules of the invention includes nucleic acids that are associated with a 15 gene or fragment thereof. Another subset of the nucleic acids of the invention includes those that encode proteins, polypeptides, or fragments of proteins or polypeptides. In one embodiment, the invention includes nucleic acid sequences having SEQ NOS. 1-82. In one embodiment, the invention includes the sequences identified in Table 1. In another embodiment, the nucleic acids of the invention 20 are derived from one or more EST sequences identified in Table 2.

Fragment nucleic acids may encompass significant portion(s) of, or indeed most of, these nucleic acids. For example, a fragment nucleic acid can encompass an OA marker gene homolog or fragment thereof. Alternatively, the fragments may comprise smaller oligonucleotides (for example, having from about 7 to 25 about 250 nucleotides, or from about 15 to about 30 nucleotide).

Nucleic acids or fragments thereof of the invention are capable of specifically hybridizing to other nucleic acids under certain circumstances. In one embodiment, a nucleic acid of the present invention will specifically hybridize to

one or more of the nucleic acids set forth in SEQ NO: 1 through SEQ NO: 82, or complements thereof, under moderately stringent conditions, for example at about 2.0 X SSC and about 65°C.

Conventional stringency conditions are described by Sambrook, *et al.*,

5 *Molecular Cloning, A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York (1989), and by Haymes, *et al. Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985). Departures from complete complementarity are therefore permissible, as long as such departures do not completely preclude the capacity of the molecules to form
10 a double-stranded structure. Thus, in order for a nucleic acid molecule to serve as a primer or probe it need only be sufficiently complementary in sequence to be able to form a stable double-stranded structure under the particular solvent and salt concentrations employed.

Appropriate stringency conditions that promote DNA hybridization, for
15 example, 6.0 X sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 X SSC at 50°C, are known to those skilled in the art or can be found in Ausubel, *et al., Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989) (see especially sections 6.3.1-6.3.6). [This reference and the supplements through December 1999 can be relied to make or use any
20 embodiment of the invention.] For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 X SSC at 50°C to a high stringency of about 0.2 X SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C. Temperature and salt
25 conditions may be varied independently.

In a particularly preferred embodiment, a nucleic acid of the invention will include those nucleic acids that specifically hybridize to one or more of the nucleic acids set forth in SEQ NO: 1 through SEQ NO: 82, or complements

thereof, under high stringency conditions. In one aspect of the present invention, the nucleic acid molecules of the present invention comprise one or more of the nucleic acid sequences set forth in SEQ NO: 1 through to SEQ NO: 82, or complements thereof.

5 In another aspect of the invention, one or more of the nucleic acid molecules of the present invention share between 80% and 100% sequence identity, or alternatively from between 90% to 100% sequence identity, with one or more of the nucleic acid sequences set forth in SEQ NO: 1 through to SEQ NO: 82 or complements thereof.

10 **(i) Nucleic Acids Comprising Genes or Fragments
Thereof**

This invention also provides genes corresponding to the cDNA sequences disclosed herein, also called OA Marker nucleic acids. The corresponding genes can be isolated in accordance with known methods using the sequence information 15 disclosed herein. The methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials.

The invention provides naturally existing gene homologues or fragments thereof. Genomic sequences can be screened for the presence of protein 20 homologues utilizing one or a number of different search algorithms that have been developed, such as the suite of BLAST programs.

In one embodiment of the present invention, the homologue protein or fragment thereof exhibits a BLASTX probability score of less than 1E-30, or alternatively a BLASTX probability score of between about 1E-30 and about 1E-25 12, or alternatively a BLASTX probability score of greater than 1E-12 with a nucleic acid or gene of this invention. In another embodiment of the present invention, the nucleic acid molecule encoding the gene homologue or fragment thereof exhibits a % identity with its homologue of between about 25% and about

99%. In a further embodiment, the gene homologue or fragment has a single nucleotide difference from its homologue.

In another preferred embodiment, nucleic acid molecules having SEQ NO: 1 through SEQ NO: 82, or complements and fragments of either, can be utilized to 5 obtain homologues equivalent to the naturally existing homologues. The degeneracy of the genetic code, which allows different nucleic acid sequences to code for the same protein or peptide, is known in the literature (see U.S. Patent No. 4,757,006). As used herein a nucleic acid molecule is degenerate of another nucleic acid molecule when the nucleic acid molecules encode for the same amino 10 acid sequences but comprise different nucleotide sequences. An aspect of the present invention is that the nucleic acid molecules of the present invention include nucleic acid molecules that are degenerate of those set forth in SEQ NO: 1 through to SEQ NO: 82 or complements thereof.

In a further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding a homologue or fragment thereof in SEQ NO: 1 through SEQ NO: 82, or complements thereof, due to the degeneracy in the genetic code in that they 5 encode the same protein but differ in nucleic acid sequence. In another further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding an homologue or fragment thereof in SEQ NO: 1 through SEQ NO: 82, or complements thereof, due to fact that the different nucleic acid sequence encodes a 10 protein having one or more conservative amino acid residue. Examples of conservative substitutions are set forth below. Codons capable of coding for such conservative

	<u>Original Residue</u>	<u>Conservative Substitutions</u>
	Ala	ser
15	Arg	lys
	Asn	gln; his
	Asp	glu
	Cys	ser; ala
	Gln	asn
20	Glu	asp
	Gly	pro
	His	asn; gln
	Ile	leu; val
	Leu	ile; val
25	Lys	arg; gln; glu
	Met	leu; ile
	Phe	met; leu; tyr
	Ser	thr

Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

5

(ii) Nucleic Acids Comprising Regulatory Elements

One class of agents of the invention includes nucleic acids having promoter regions or partial promoter regions or regulatory elements. Promoter regions are typically found upstream of the trinucleotide ATG sequence at the start 10 site of a protein coding region. The nucleic acids of the invention may be used to isolate promoters of cell-enhanced, cell-specific, tissue-enhanced, tissue-specific, developmentally- or physiologically-regulated expression profiles. Isolation and functional analysis of the 5' flanking promoter sequences from genomic libraries, for example, using genomic screening methods and PCR techniques, results in the 15 isolation of useful promoters and transcriptional regulatory elements. These methods are known to those of skill in the art and have been described (see, for example, Birren *et al.*, *Genome Analysis: Analyzing DNA*, 1, (1997), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.).

For example, in one embodiment, a regulatory element is detected by 20 incubating nucleic acid(s), or preferably fragments such as ESTs, with members of genomic libraries (e.g., synovial tissue libraries) and recovering clones that hybridize to the nucleic acid(s). Sequencing techniques can then identify regulatory elements from known sequence motifs or known assays for detecting regulatory sequences within a certain proximity to transcription and translation 25 start and stop sites can be used. In a second embodiment, methods of "chromosome walking," or inverse PCR may be used to obtain regulatory elements (Frohman *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 85:8998-9002 (1988); Ohara *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 86: 5673-5677 (1989); Pang *et al.*,

Biotechniques 22(6): 1046-1048 (1997); Huang *et al.*, *Methods Mol. Biol.* 69: 89-96 (1997); Huang, *et al.*, *Method Mol. Biol.* 67:287-294 (1997); Benkel *et al.*, *Genet. Anal.* 13: 123-127 (1996); Hartl *et al.*, *Methods Mol. Biol.* 58: 293-301 (1996)).

5 Promoters and regulatory elements obtained utilizing the nucleic acids of the invention can also be modified to affect their control characteristics. Examples of these modifications include, but are not limited to, enhancer sequences as reported by Kay *et al.*, *Science* 236:1299 (1987). Genetic elements such as these can be used to enhance gene expression of new and existing proteins
10 or polypeptides.

(b) Proteins and Polypeptides

This invention also provides a compound or composition comprising one or more polypeptides, which comprise: 1) at least one fragment, segment, or domain of at least 15-1,000 contiguous amino acids, with at least one portion
15 encoded by one or more of SEQ NOS: 1-82; 2) at least one amino acid sequence selected from those encoded by at least one of SEQ NOS: 1-82; or 3) at least one modification corresponding to fragments, segments, or domains within one of SEQ NOS:1- 82.

Protein and peptide molecules can be identified using known protein or
20 peptide molecules as a target sequence or target motif in the BLAST programs of the present invention. Proteins or peptides may undergo a variety of modifications, including post-translational modifications, such as disulfide bond formation, glycosylation, phosphorylation, or oligomerization. The term "protein" or "polypeptide" includes any protein molecule that is modified by any biological
25 or non-biological process. Proteins encoded by sequences of the present invention can also be fusion proteins. A fusion protein or peptide molecule of the present invention is preferably produced via recombinant means.

Another class of agents comprises protein or peptide molecules encoded by SEQ NO: 1 through SEQ NO: 82 or complements thereof or, fragments or fusions thereof in which conservative, non-essential, or not relevant, amino acid residues have been added, replaced, or deleted. An example is the homologue protein of an

5 OA marker protein. Such a homologue can be obtained by any of a variety of methods. For example, as indicated above, one or more of the disclosed sequences (SEQ NO: 1 through SEQ NO: 82, or complements thereof) will be used to define a pair of primers that may be used to isolate the homologue-encoding nucleic acid molecules from any desired species. Such molecules can be

10 expressed to yield homologs by recombinant means.

Proteins or polypeptides of the invention can be expressed as variants that facilitate purification. For example, a fusion protein to such proteins as maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX) are known in the art [New England BioLab, Beverly, Mass., Pharmacia, Piscataway, 15 N.J., and InVitrogen, San Diego, CA]. The polypeptide or protein can also be a tagged variant to facilitate purification, such as with histidine or methionine rich regions [His-Tag; available from LifeTechnologies Inc, Gaithersburg, MD] that bind to metal ion affinity chromatography columns, or with an epitope that binds to a specific antibody [Flag, available from Kodak, New Haven, Conn.]. An 20 exemplary, non-limiting list of commercially available vectors suitable for fusion protein expression includes: pBR322 (Promega); pGEX (Amersham); pT7 (USB); pET (Novagen); pIBI (IBI); pProEX-1 (Gibco/BRL); pBluescript II (Stratagene); pTZ18R and pTZ19R (USB); pSE420 (Invitrogen); pAc360 (Invitrogen); pBlueBac (Invitrogen); pBacPAK (Clontech); pHIL (Invitrogen); pYES2 25 (Invitrogen); pCDNA (Invitrogen); and pREP (Invitrogen).

A number of other purification methods or means are also known and can be used. These include, but are not limited to, anion or cation exchange chromatography, ethanol precipitation, affinity chromatography, high performance

liquid chromatography (HPLC), and reverse-phase high performance liquid chromatography (RP-HPLC), optionally employing hydrophobic RP-HPLC media, e.g., silica gel, to further purify the protein. Combinations of methods and means can also be employed to provide a substantially purified recombinant polypeptide or protein. The particular method used will depend upon the properties of the polypeptide, and the selection of the host cell; appropriate methods will be readily apparent to those skilled in the art.

The polypeptide or protein of the invention may also be expressed via transgenic animals. Methods and means employing the milk of transgenic 10 domestic animals are known in the art.

One or more of the proteins, polypeptides, or fragments may be produced via chemical synthesis. Methods for synthetic construction are known to those skilled in the art. The synthetically-constructed sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with 15 proteins, may possess biological properties in common, including protein activity. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

(c) Antibodies

20 One aspect of the present invention concerns antibodies, single-chain antigen binding molecules, or other proteins that specifically bind to one or more of the protein or peptide molecules of the present invention and their homologues, fusions or fragments. Such antibodies may be used to quantitatively or qualitatively detect the protein or peptide molecules of the present invention. As 25 used herein, an antibody or peptide is said to "specifically bind" to a protein or peptide molecule of the present invention if such binding is not competitively inhibited by the presence of non-related molecules.

Nucleic acid molecules that encode all or part of the protein of the present invention can be expressed, by recombinant means, to yield protein or peptides that can in turn be used to elicit antibodies that are capable of binding the expressed protein or peptide. Such antibodies may be used in immunoassays for 5 that protein. Such protein-encoding molecules or their fragments may be a "fusion" molecule (*i.e.*, a part of a larger nucleic acid molecule) such that, upon expression, a fusion protein is produced. It is understood that any of the nucleic acid molecules of the present invention may be expressed, by recombinant means, to yield proteins or peptides encoded by these nucleic acid molecules.

10 The antibodies that specifically bind proteins and protein fragments of the present invention may be polyclonal or monoclonal, and may comprise intact immunoglobulins, or antigen binding portions of immunoglobulins (such as (F(ab')₁, F(ab')₂ fragments), or single-chain immunoglobulins producible, for example, via recombinant means. Conditions and procedures for the construction, 15 manipulation and isolation of antibodies (see, for example, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, New York (1988)) are well known in the art.

Additionally, methods of raising antibodies to polypeptides of the present invention are well known in the art (Harlow and Lane, *Antibodies: A Laboratory 20 Manual*, Cold Spring Harbor Press, Cold Spring Harbor, New York (1988)). Animals are immunized with purified protein (or fragment thereof) in appropriate amounts and using standard schedules, and are bled and the blood screened for the presence of anti-protein or peptide antibodies using, for example, a direct binding Enzyme-Linked Immunoassay (ELISA). In one embodiment, anti-protein or 25 peptide monoclonal antibodies are isolated using a fusion of a protein, protein fragment, or peptide of the present invention, or conjugate of a protein, protein fragment, or peptide of the present invention, as immunogens. It is understood

that any of the protein or peptide molecules of the present invention may be used to raise antibodies.

As discussed below, such antibody molecules or their fragments may be used for diagnostic purposes. Where the antibodies are intended for diagnostic purposes, it may be desirable to derivatize them, for example with a ligand group (such as biotin) or a detectable marker group (such as a fluorescent group, a radioisotope or an enzyme).

The ability to produce antibodies that bind the protein or peptide molecules of the present invention permits the identification of mimetic compounds of those molecules. Mimetic compounds can be synthesized chemically. Combinatorial chemistry techniques, for example, can be used to produce libraries of peptides (see WO 9700267), polyketides (see WO 960968), peptide analogues (see WO 9635781, WO 9635122, and WO 9640732), oligonucleotides for use as mimetic compounds derived from this invention. Mimetic compounds and libraries can also be generated through recombinant DNA-derived techniques. For example, phage display libraries (see WO 9709436), DNA shuffling (see US Patent 5,811,238) other directed or random mutagenesis techniques can produce libraries of expressed mimetic compounds. It is understood that any of the agents of the present invention can be substantially purified and/or be biologically active and/or recombinant.

(d) Transformed Cells

A recombinant protein may be produced by operably linking a regulatory control sequence to a nucleic acid of the present invention and putting it into an expression vector. Regulatory sequences include promoters, enhancers, and other expression control elements which are described in Goeddel (*Heter Expression Technology: Methods in Enzymology* 185. Academic Press, San Diego, CA (1990)). For example, the native regulatory sequences or regulatory sequences native to the transformed host cell can be used. One of skill in the art is familiar

with numerous examples of these additional functional sequences, as well as other functional sequences, that may optionally be included in an expression vector. The design of the expression vector may depend on such factors as the choice of the host cell to be transformed, and/or the type of protein desired. Many such 5 vectors are commercially available, including linear or enclosed elements (see for example, Broach, et al., *Experimental Manipulation of Gene Expression*, ed. M. Inouye, Academic Press, (1983); Sambrook, et al., *Molecular Cloning, A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York (1989)). Typically, expression constructs will contain one or more 10 selectable markers, including the gene that encodes dihydrofolate reductase and the genes that confer resistance to neomycin, tetracycline, ampicillin, chloramphenicol, kanamycin and streptomycin resistance.

Prokaryotic and eukaryotic host cells transfected by the described vectors are also provided by this invention. For instance, cells which can be transfected 15 with the vectors of the present invention include, but are not limited to, bacterial cells such as *E. coli* (e.g., *E. coli* K 12 strains), *Streptomyces*, *Pseudomonas*, *Serratia marcescens* and *Salmonella typhimurium*, insect cells (baculovirus), including *Drosophila*, fungal cells, such as yeast cells, plant cells, and ovary cells (CHO), and COS cells.

20 Thus, the nucleic acid molecules described can be used to produce a recombinant form of the protein via microbial or eukaryotic cellular processes. Ligating the polynucleic acid molecule into a gene construct, such as an expression vector, and transforming or transfecting into hosts, either eukaryotic (yeast, avian, insect, plant, or mammalian) or prokaryotic (bacterial cells), are 25 standard procedures used in producing other well known proteins. Similar procedures, or modifications thereof, can be employed to prepare recombinant proteins according to the present invention by microbial means or tissue-culture

technology. Accordingly, the invention pertains to the production of encoded proteins or polypeptides by recombinant technologies.

D. Uses of the Agents of the Invention

1. Methods for Identifying Bioactive Proteins, Polypeptides, Fragments, or Variants of the Invention

Once the nucleic acid has been used to produce a protein, polypeptide, or a variant or fragment, any one of a number of assays can be used to identify bioactivity.

In addition, the agents of the invention are especially useful in high throughput screening methods. In general, these methods involve individual sample assay volumes less than about 250 μ l. or more preferably less than about 100 μ l. With smaller sample volumes, numerous individual assays can be performed simultaneously and via computer-operated instrumentation. The assays comprise the detectable interaction between a protein, polypeptide, fragment, nucleic acid, or antibody of the invention (sometimes referred to as the target) and an assay compound. Thus, the assays comprise two components, a target and an assay compound, where the assay compound may be part of a composition of multiple compounds. It is also possible for the agents of this invention to be used as assay compounds in screening methods where other proteins, polypeptides, nucleic acids, antibodies, or binding partners are the targets.

The assay compound can be selected from a library of small molecules, organic compounds which are either synthetic or natural, or mimetic libraries of randomized oligonucleotide-derived or peptide-derived compounds, for example. The compounds of the libraries may contain random chemical modifications, such as acylation, alkylation, esterification, amidation, or other modifications. Ideally, the largest number of separate structural entities will exist in a library that is tested against the agents of the invention for detectable interaction. A variety of other

reagents may be used in the assay, such as buffers, salts, detergents, proteins, protease inhibitors, nuclease inhibitors, antimicrobial agents, or other reagents.

Detecting the interaction between the assay compound and the agent of the invention can be performed via a number of techniques. Fluorescence quenching, 5 specific binding as with avidin-biotin, enzymatic activity, or inhibition of enzymatic activity are examples of the types of techniques used to detect interaction between two molecules. One of skill in the art can devise many specific assays depending on the activity sought. The type of assay used is not crucial to the use of this invention.

10 The screening methods may optionally employ a solid substrate to which one or more assay components are bound. Also, cell-based assays are often used in high throughput screening methods, so that the cell contains or expresses a component of the assay. Numerous permutations are possible.

15 Each of the activities listed below may be screened for, alone or in combination, in a method to detect an interaction with agents of the invention.

a. Inflammatory or Anti-Inflammatory Activity

Proteins or polypeptides of the present invention may also exhibit inflammatory or anti-inflammatory activity. These activities may relate to a stimulus to cells involved in the inflammatory response, inhibiting or promoting 20 cell-cell interactions (for example, cell adhesion), inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or stimulating or suppressing production of other factors, which more directly inhibit or promote an inflammatory response.

Proteins or polypeptides exhibiting anti-inflammatory activity or 25 antibodies to inflammatory proteins or polypeptides can be used to treat atopic disorders and other inflammatory conditions including: chronic or acute inflammatory conditions, inflammation associated with infection, such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS); ischemia-

reperfusion injury; endotoxin lethality; arthritis; complement-mediated hyperacute rejection; nephritis; cytokine or chemokine-induced lung injury; inflammatory bowel disease; Crohn's disease; or disorders resulting from over-production of cytokines such as TNF or IL-1. Proteins or polypeptides or antibodies of the 5 invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

b. Cytokine and Cell Growth or Differentiation Activity

A protein or polypeptide of the invention may exhibit cytokine, cell growth promoting or inhibiting, or cell differentiation promoting or inhibiting 10 activity. Many protein factors secreted by immune cells, including cytokines, have exhibited activity in one or more factor dependent cell-based assays. These assays can be used to identify useful activities. The activity of a protein or polypeptide of the invention may be measured by the following methods or others known in the art.

15 Assays for T-cell or thymocyte proliferation include those described in *Current Protocols in Immunology*, Coligan, *et al.* Eds., Greene Publishing Associates and Wiley-Interscience (1994) (see, especially Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Takai *et al.*, *J. Immunol.* 137:3494-3500 (1986); Bertagnolli *et al.*, *J. 20 Immunol.* 145:1706-1712 (1990); Bertagnolli *et al.*, *Cellular Immunology* 133:327-341 (1991); Bertagnolli, *et al.*, *J. Immunol.* 149:3778-3783 (1992); Bowman *et al.*, *J. Immunol.* 152: 1756-1761 (1994)).

Assays for cytokine production and/or proliferation of spleen cells, lymph 25 node cells, or thymocytes include those described in Polyclonal T cell stimulation, Kruisbeek, and Shevach, In *Current Protocols in Immunology*, Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto (1994); and *Measurement of mouse and human Interferon gamma*, Schreiber, In *Current Protocols in*

Immunology, Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto (1994).

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include those described in *Measurement of Human and Murine Interleukin 2 and Interleukin 4*, Bottomly, Davis, and Lipsky In *Current Protocols in Immunology*. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto (1994); deVries *et al.*, *J. Exp. Med.* 173:1205-1211 (1991); Moreau *et al.*, *Nature* 336:690-692 (1988); Greenberger *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 80:2931-2938 (1983); *Measurement of mouse and human interleukin 6*, Nordan, R. In *Current Protocols in Immunology*. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto (1994); Smith *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 83:1857-1861 (1986); *Measurement of human Interleukin 11*, Bennett, *et al.*, In *Current Protocols in Immunology*. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto (1991); *Measurement of mouse and human Interleukin 9*, Ciarletta, *et al.*, In *Current Protocols in Immunology*. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto (1991).

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect antigen-presenting cell (APC)-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: *Current Protocols in Immunology*, Coligan, *et al.* eds.. Pub. Greene Publishing Associates and Wiley-Interscience (1994)(Chapter 3, *In Vitro assays for Mouse Lymphocyte Function*; Chapter 6, *Cytokines and their cellular receptors*; Chapter 7, *Immunologic studies in Humans*); Weinberger *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 77:6091-6095, 1980; Weinberger *et al.*, *Eur. J. Immun.* 11:405-411, 1981; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512 (1988).

-32-

**c. Immunosuppressive, Immune Stimulating, or
Immune Modulating Activity**

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: *Current Protocols in Immunology*, Ed by Coligan, *et al.*, Pub. Greene Publishing Associates and Wiley-Interscience (1994)(Chapter 3, *In Vitro assays for Mouse Lymphocyte Function* 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Herrmann *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 78:2488-2492 (1981); Herrmann *et al.*, *J. Immunol.* 128:1968-1974 (1982); Handa *et al.*, *J. Immunol.* 135:1564-1572 (1985); Takai *et al.*, *J. Immunol.* 137:3494-3500 (1986); Takai *et al.*, *J. Immunol.* 140:508-512 (1988); Bowman *et al.*, *J. Virology* 61:1992-1998; Bertagnolli *et al.*, *Cellular Immunology* 133:327-341, 1991; Brown *et al.*, *J. Immunol.* 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, *J. Immunol.* 144:3028-3033 (1990); and *Assays for B cell function: In vitro antibody production*, Mond, and Brunswick, In *Current Protocols in Immunology*, Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto (1994).

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: *Current Protocols in Immunology*, Ed by Coligan, *et al.*, Strober, Pub. Greene Publishing Associates and Wiley-Interscience (1994) (Chapter 3, *In Vitro assays for Mouse Lymphocyte Function* 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988; Bertagnolli *et al.*, *J. Immunol.* 149:3778-3783 (1992).

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without

limitation, those described in: Guery *et al.*, *J. Immunol.* 134:536-544, 1995; Inaba *et al.*, *Journal of Experimental Medicine* 173:549-559, 1991; Macatonia *et al.*, *J Immunol* 154:5071-5079 (1995); Porgador *et al.*, *Journal of Experimental Medicine* 182:255-260 (1995); Nair *et al.*, *J. Virology* 67:4062-4069 (1993);
5 Huang *et al.*, *Science* 264:961-965 (1994); Macatonia *et al.*, *Journal of Experimental Medicine* 169:1255-1264 (1989); Bhardwaj *et al.*, *Journal of Clinical Investigation* 94:797-807 (1994); and Inaba *et al.*, *Journal of Experimental Medicine* 172:631-640 (1990).

Assays for lymphocyte survival/apoptosis (which will identify, among
10 others, proteins that prevent apoptosis after superantigen induction and proteins
that regulate lymphocyte homeostasis) include, without limitation, those described
in: Darzynkiewicz *et al.*, *Cytometry* 13:795-808 (1992); Gorczyca *et al.*, *Leukemia*
7:659-670 (1993); Gorczyca *et al.*, *Cancer Research* 53:1945-1951 (1993); Itoh *et
al.*, *Cell* 66:233-243 (1991); Zacharchuk, *J. Immunol.* 145:4037-4045 (1990);
15 Zamai *et al.*, *Cytometry* 14:891-897 (1993); Gorczyca *et al.*, *International Journal
of Oncology* 1:639-648 (1992).

Assays for proteins that influence early steps of T-cell commitment and
development include, without limitation, those described in: Antica *et al.*, *Blood*
84:111-117 (1994); Fine *et al.*, *Cellular Immunology* 155:111-122 (1994); Galy *et
al.*, *Blood* 85:2770-2778 (1995); Toki *et al.*, *Proc. Nat. Acad. Sci. (U.S.A.)*
88:7548-7551 (1991).

d. Cell Differentiation Activity

Assays for embryonic stem cell differentiation (which will identify, among
others, proteins that influence embryonic differentiation hematopoiesis) include,
25 without limitation, those described in: Johansson *et al.*, *Cellular Biology* 15:141-
151 (1995); Keller *et al.*, *Molecular and Cellular Biology* 13:473-486 (1993);
McClanahan *et al.*, *81:2903-2915 (1993).*

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: *Methylcellulose colony forming assays*, Freshney, In *Culture of Hematopoietic Cells*, Freshney, et al., eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. (1994); Hirayama et al., *Proc. Natl. Acad. Sci. (U.S.A.)* 89:5907-5911 (1992); Primitive hematopoietic colony forming cells with high proliferate potential, McNiece, and Briddell, In *Culture of Hematopoietic Cells*, Freshney, et al., eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. (1994); Neben et al., *Experimental Hematology* 22:353-359 (1994); Cobblestone area forming cell assay, Ploemacher, In *Culture of Hematopoietic Cells*, Freshney, et al., eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. (1994); and Long term bone marrow cultures in the presence of stromal cells, Spooncer, et al., In *Culture of Hematopoietic Cells*, Freshney, et al., eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. (1994); Long term culture initiating cell assay, Sutherland, In *Culture of Hematopoietic Cells*, R. I. Freshney, et al., eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. (1994).

e. Wound Healing Activity

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); and International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, *Epidermal Wound Healing*. pps. 71-112 (Maibach, and Rovee, eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, *J. Invest. Dermatol* 71:382-84 (1978).

f. Chemotactic Activity

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the controlled orientation or

movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

5 The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein
10 to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: *Current Protocols in Immunology*, Ed by Coligan *et al.*, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub *et al.*, *J. Clin. Invest.* 95:1370-1376 (1995);
15 Lind *et al.*, *APMIS* 103:140-146 (1995); Muller *et al.*, *Eur. J. Immunol.* 25: 1744-1748; Gruber *et al.*, *J. of Immunol.* 152:5860-5867 (1994); and Johnston *et al.* *J. of Immunol.* 153: 1762-1768 (1994).

g. Receptor/Ligand Interaction Activity

Proteins of the present invention may also demonstrate activity as
20 receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as
25 selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). These proteins or fragments of the invention, or cells containing them, can be incorporated into an assay to screen for binding to

extracellular matrix proteins, their analogs, or for receptor/ligand interaction to compounds implicated in binding to extracellular matrix proteins. In a particularly preferred embodiment, molecules containing the peptide motif RGD can be used to screen for interaction with the proteins, fragments, or cells of the 5 invention. Various specific assays can be used a basis for designing the reagents for screening, such as phage attachment assays, panning assays, cell attachment assays, and inhibition of cell attachment/adhesion assays (Pasquelina *et al.*, *J. Cell Biol.* 130:1189-1196 (1995), Koivunen *et al.*, *BioTechnology* 13: 265-270 (1995), Koivunen *et al.*, *Methods Enzym.* 245: 346-369 (1994), and U.S. Patent 5,817,750, 10 all incorporated herein in their entirety). These receptor/ligand interaction assays can also be designed for use with libraries of compounds, such as phage display libraries.

Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. Proteins of 15 the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those 20 described in: *Current Protocols in Immunology*, Ed by Coligan, *et al.*, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under Static Conditions 7.28.1-7.28.22); Takai *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 84:6864-6868 (1987); Bierer *et al.*, *J. Exp. Med.* 168:1145-1156 (1988); Rosenstein *et al.*, *J. Exp. Med.* 169:149-160 (1989); 25 Stoltenborg *et al.*, *J. Immunol. Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670 (1995).

2. Methods for Detecting and Manipulating Nucleic Acids

The nucleic acids of the invention can be used directly in numerous methods to identify or detect the presence of specific nucleic acid sequences. As noted above, the nucleic acids of the invention can be used as hybridization probes or PCR probes, or to derive specific hybridization or PCR probes. Agents of the present invention may be labeled with reagents that facilitate detection (e.g., fluorescent labels, Prober *et al.*, *Science* 238: 336-340 (1987), Albarella *et al.*, EP 144914; chemical labels, Sheldon *et al.*, U.S. Patent 4,582,789, Albarella *et al.*, U.S. Patent 4,563,417; and modified bases, Miyoshi *et al.*, EP 119448)).

Furthermore, the nucleic acids of the invention of variants or fragments thereof can be linked to solid supports. In this way, various microarrays, beads, glass or nylon slides, membranes or other repeatable assay apparati can be constructed. A non-limiting description of selected methods follows.

a. Microarrays

In one embodiment, the nucleic acids of the invention can be used to monitor expression. A microarray-based method for high-throughput monitoring of gene expression may be utilized to measure activated OA Marker hybridization targets. This 'chip'-based approach involves using microarrays of nucleic acids as specific hybridization targets to quantitatively measure expression of the corresponding genes (Schena *et al.*, *Science* 270:467-470 (1995); Shalon, Ph.D. Thesis, Stanford University (1996). Every nucleotide in a large sequence can be queried at the same time. Hybridization can also be used to efficiently analyze nucleotide sequences.

Several microarray methods have been described. One method compares the sequences to be analyzed by hybridization to a set of oligonucleotides or cDNA molecules representing all possible subsequences (Bains and Smith, *J. Theor. Biol.* 135:303 (1989)). A second method hybridizes the sample to an array of oligonucleotide or cDNA probes. An array consisting of oligonucleotides or

cDNA molecules complementary to subsequences of a target sequence can be used to determine the identity of a target sequence, measure its amount, and detect differences between the target and a reference sequence. Nucleic acid microarrays may also be screened with protein molecules or fragments thereof to determine 5 nucleic acids that specifically bind protein molecules or fragments thereof.

The microarray approach may also be used with polypeptide targets (see, U.S. Patent Nos. 5,800,992, 5,445,934; 5,143,854, 5,079,600, 4,923,901). Essentially, polypeptides are synthesized on a substrate (microarray) and these polypeptides can be screened with either protein molecules or fragments thereof or 10 nucleic acid molecules in order to screen for either protein molecules or fragments thereof or nucleic acid molecules that specifically bind the target polypeptides (Fodor *et al.*, *Science* 251:767-773 (1991)).

b. Hybridization Assays

Oligonucleotide probes, whose sequences are complementary to that of a 15 portion of the nucleic acids of the invention, such as SEQ NO.:1-82, can be constructed. These probes are then incubated with cell extracts of a patient under conditions sufficient to permit nucleic acid hybridization. The detection of double-stranded probe-mRNA hybrid molecules is indicative of the presence of OA Markers or sequences derived from OA Markers. Thus, such probes may be 20 used to ascertain the level and extent of OA severity or progression or the production of certain proteins. The nucleic acid hybridization may be conducted under quantitative conditions or as a qualitative assay.

c. PCR Assays

A nucleic acid of the invention, such as one of SEQ NO.:1-82 or 25 complements thereof, can be analyzed for use as a PCR probe. A search of databases indicates the presence of regions within that nucleic acid that have high and low regions of identity to other sequences in the database. Ideally, a PCR probe will have high identity with only the sequence from which it is derived. In

that way, only the desired sequence is amplified. Computer generated searches using programs such as MIT Primer3 (Rozen and Skaletsky (1996, 1997, 1998)), or GeneUp (Pesole, *et al.*, *BioTechniques* 25:112-123 (1998)), for example, can be used to identify potential PCR primers.

5 The PCR probes or primers can be used in methods such as described in Krzesicki, *et al.*, *Am. J. Respir. Cell Mol. Biol.* 16:693-701 (1997) to identify or detect sequences expressed in osteoarthritis.

d. Ligation and Alternative Amplification Methods and Identification of Polymorphisms

10 In one sub-aspect of the invention analysis is conducted to determine the presence and/or identity of polymorphism(s) using one or more of the nucleic acid molecules of the present invention and more specifically one or more of the EST nucleic acid molecule or fragment thereof which are associated with a phenotype, or a predisposition to that phenotype.

15 Any of a variety of molecules can be used to identify such polymorphism(s). In one embodiment, one or more of the EST nucleic acid molecules (or a sub-fragment thereof) may be employed as a marker nucleic acid molecule to identify such polymorphism(s). Alternatively, such polymorphisms can be detected through the use of a marker nucleic acid molecule or a marker 20 protein that is genetically linked to (i.e., a polynucleotide that co-segregates with) such polymorphism(s).

25 In an alternative embodiment, such polymorphisms can be detected through the use of a marker nucleic acid molecule that is physically linked to such polymorphism(s). For this purpose, marker nucleic acid molecules comprising a nucleotide sequence of a polynucleotide located within 1 mb of the polymorphism(s), and more preferably within 100 kb of the polymorphism(s), and most preferably within 10 kb of the polymorphism(s) can be employed.

The genomes of animals and plants naturally undergo spontaneous mutation in the course of their continuing evolution, creating polymorphisms (Gusella, *Ann. Rev. Biochem.* 55:831-854 (1986)). The variant sequence and the "original" sequence co-exist in the species' population. In some instances, such co-existence is in 5 stable or quasi-stable equilibrium.

A polymorphism is thus said to be "allelic," in that, due to the existence of the polymorphism, some members of a species may have the original sequence (i.e., the original "allele") whereas other members may have the variant sequence (i.e., the variant "allele"). In the simplest case, only one variant sequence may 10 exist, and the polymorphism is thus said to be di-allelic. In other cases, the species' population may contain multiple alleles, and the polymorphism is termed tri-allelic, etc. A single gene may have multiple different unrelated polymorphisms. For example, it may have a di-allelic polymorphism at one site, and a multi-allelic polymorphism at another site.

15 The variation that defines the polymorphism may range from a single nucleotide variation to the insertion or deletion of extended regions within a gene. In some cases, the DNA sequence variations are in regions of the genome that are characterized by short tandem repeats (STRs) that include tandem di- or tri-nucleotide repeated motifs of nucleotides. Polymorphisms characterized by such 20 tandem repeats are referred to as "variable number tandem repeat" ("VNTR") polymorphisms. VNTRs have been used in identity analysis (Weber, U.S. Patent 5,075,217; Armour, *et al.*, *FEBS Lett.* 307:113-115 (1992); Jones, *et al.*, *Eur. J. Haematol.* 39:144-147 (1987); Horn, *et al.*, PCT Patent Application WO91/14003; Jeffreys, European Patent Application 370,719; Jeffreys, U.S. Patent 5,175,082; 25 Jeffreys *et al.*, *Amer. J. Hum. Genet.* 39:11-24 (1986); Jeffreys *et al.*, *Nature* 316:76-79 (1985); Gray *et al.*, *Proc. R. Acad. Soc. Lond.* 243:241-253 (1991); Moore *et al.*, *Genomics* 10:654-660 (1991); Jeffreys *et al.*, *Anim. Genet.* 18:1-15

(1987); Hillel *et al.*, *Anim. Genet.* 20:145-155 (1989); Hillel *et al.*, *Genet.* 124:783-789 (1990)).

The detection of polymorphic sites in a sample of DNA may be facilitated through the use of nucleic acid amplification methods. Such methods specifically 5 increase the concentration of polynucleotides that span the polymorphic site, or include that site and sequences located either distal or proximal to it. Such amplified molecules can be readily detected by gel electrophoresis or other means.

The most preferred method of achieving such amplification employs the polymerase chain reaction ("PCR") (See above). In lieu of PCR, alternative 10 amplification methods, such as the "Ligase Chain Reaction" ("LCR") may be used (Barany, *Proc. Natl. Acad. Sci. (U.S.A.)* 88:189-193 (1991)). LCR uses two pairs of oligonucleotide probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such 15 hybridization forms a substrate for a template-dependent ligase. As with PCR, the resulting products thus serve as a template in subsequent cycles and an exponential amplification of the desired sequence is obtained.

LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a polymorphic site. In one embodiment, 20 either oligonucleotide will be designed to include the actual polymorphic site of the polymorphism. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the polymorphic site present on the oligonucleotide. Alternatively, the 25 oligonucleotides may be selected such that they do not include the polymorphic site (see, Segev, PCT Application WO 90/01069).

The "Oligonucleotide Ligation Assay" ("OLA") may alternatively be employed (Landegren, *et al.*, *Science* 241:1077-1080 (1988)). The OLA protocol

uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target. OLA, like LCR, is particularly suited for the detection of point mutations. Unlike LCR, however, OLA results in "linear" rather than exponential amplification of the target sequence.

5 Nickerson, *et al.* have described a nucleic acid detection assay that combines attributes of PCR and OLA (Nickerson, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 87:8923-8927 (1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA. In addition to requiring multiple, and separate, processing steps, one problem 10 associated with such combinations is that they inherit all of the problems associated with PCR and OLA.

Schemes based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide," thereby amplifying the di-oligonucleotide, are also known (Wu, *et al.*, *Genomics* 15 4:560 (1989)), and may be readily adapted to the purposes of the present invention.

Other known nucleic acid amplification procedures, such as allele-specific oligomers, branched DNA technology, transcription-based amplification systems, or isothermal amplification methods may also be used to amplify and analyze such 20 polymorphisms (Malek, *et al.*, U.S. Patent 5,130,238; Davey, *et al.*, European Patent Application 329,822; Schuster *et al.*, U.S. Patent 5,169,766; Miller, *et al.*, PCT appln. WO 89/06700; Kwoh, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 86:1173 (1989); Gingeras, *et al.*, PCT application WO 88/10315; Walker, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 89:392-396 (1992)). Any of the foregoing nucleic acid 25 amplification methods could be used.

The identification of a polymorphism in a gene, fragment or cellular sequence derived from the nucleic acids of the invention can be determined in a variety of ways. By correlating the presence or absence of atopic disease, for

example, in an individual with the presence or absence of a polymorphism, it is possible to diagnose the predisposition of a patient to osteoarthritis-related disorders. If a polymorphism creates or destroys a restriction endonuclease cleavage site, or if it results in the loss or insertion of DNA (e.g., a VNTR polymorphism), it will alter the size or profile of the DNA fragments that are generated by digestion with that restriction endonuclease. As such, individuals that possess a variant sequence can be distinguished from those having the original sequence by restriction fragment analysis. Polymorphisms that can be identified in this manner are termed "restriction fragment length polymorphisms" ("RFLPs"). RFLPs have been widely used in human and animal genetic analyses (Glassberg, UK Patent Application 2135774; Skolnick, *et al.*, *Cytogen. Cell Genet.* 32:58-67 (1982); Botstein, *et al.*, *Ann. J. Hum. Genet.* 32:314-331 (1980); Fischer, *et al.* (PCT Application WO90/13668); Uhlen, PCT Application WO90/11369).

Other types of polymorphisms include single nucleotide polymorphisms (SNPs) that are single base changes in genomic DNA sequence. They generally occur at greater frequency than other markers and are spaced with a greater uniformity throughout a genome than other reported forms of polymorphism. The greater frequency and uniformity of SNPs means that there is greater probability that such a polymorphism will be found near or in a genetic locus of interest than would be the case for other polymorphisms. SNPs are located in protein-coding regions and noncoding regions of a genome. Some of these SNPs may result in defective or variant protein expression (e.g., as a result of mutations or defective splicing). Analysis (genotyping) of characterized SNPs can require only a plus/minus assay rather than a lengthy measurement, permitting easier automation.

Modifications can be naturally provided or deliberately engineered into the nucleic acids, proteins, and polypeptides of the invention to generate variants. For example, modifications in the peptide or DNA sequences can be made by those

skilled in the art using known techniques, such as site-directed mutagenesis. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of one or more selected amino acid residues. For example, one or more cysteine residues may be deleted or replaced 5 with another amino acid to alter the conformation of the molecule. Additional cysteine residues can also be added as a substitute at sites to promote disulfide bonding and increase stability. Techniques for identifying the sites for alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art. Techniques for making alterations, substitutions, replacements, insertions 10 or deletions (see, e.g., U.S. Pat. No. 4,518,584) are also well known in the art. Preferably, any modification of a protein, polypeptide, or nucleic acid of the invention will retain at least one of the structural or functional attributes of the molecule.

SNPs can be characterized using any of a variety of methods. Such 15 methods include the direct or indirect sequencing of the site, the use of restriction enzymes (Botstein *et al.*, *Am. J. Hum. Genet.* 32:314-331 (1980)); Konieczny and Ausubel, *Plant J.* 4:403-410 (1993), enzymatic and chemical mismatch assays (Myers *et al.*, *Nature* 313:495-498 (1985), allele-specific PCR (Newton *et al.*, *Nucl. Acids Res.* 17:2503-2516 (1989)); Wu *et al.*, *Proc. Natl. Acad. Sci. USA* 20 86:2757-2760 (1989)), ligase chain reaction (Barany, *Proc. Natl. Acad. Sci. USA* 88:189-193 (1991)), single-strand conformation polymorphism analysis (Labrune *et al.*, *Am. J. Hum. Genet.* 48: 1115-1120 (1991)), primer-directed nucleotide incorporation assays (Kuppuswami *et al.*, *Proc. Natl. Acad. Sci. USA* 88:1143-1147 (1991)), dideoxy fingerprinting (Sarkar *et al.*, *Genomics* 13:441-443 (1992)), 25 solid-phase ELISA-based oligonucleotide ligation assays (Nikiforov *et al.*, *Nucl. Acids Res.* 22:4167-4175 (1994)), oligonucleotide fluorescence-quenching assays (Livak *et al.*, *PCR Methods Appl.* 4:357-362 (1995a)), 5'-nuclease allele-specific hybridization TaqMan assay (Livak *et al.*, *Nature Genet.* 9:341-342 (1995)),

template-directed dye-terminator incorporation (TDI) assay (Chen and Kwok, *Nucl. Acids Res.* 25:347-353 (1997)), allele-specific molecular beacon assay (Tyagi *et al.*, *Nature Biotech.* 16: 49-53 (1998)), PinPoint assay (Haff and Smirnov, *Genome Res.* 7: 378-388 (1997)), and dCAPS analysis (Neff *et al.*, *Plant J.* 14:387-392 (1998)).

SNPs can be observed by examining sequences of overlapping clones in the BAC library according to the method described by Taillon-Miller *et al.* *Genome Res.* 8:748-754 (1998)). SNPs can also be observed by screening the BAC library of the present invention by colony or plaque hybridization with a 10 labeled probe containing SNP markers; isolating positive clones and sequencing the inserts of the positive clones; suitable primers flanking the SNP markers.

Polymorphisms can also be identified by Single Strand Conformation Polymorphism (SSCP) analysis. SSCP is a method capable of identifying most sequence variations in a single strand of DNA, typically between 150 and 250 15 nucleotides in length (Elles, *Methods in Molecular Medicine: Molecular Diagnosis of Genetic Diseases*, Humana Press (1996); Orita *et al.*, *Genomics* 5:874-879 (1989)). Under denaturing conditions a single strand of DNA will adopt a conformation that is uniquely dependent on its sequence conformation. This conformation usually will be different, even if only a single base is changed. 20 Most conformations have been reported to alter the physical configuration or size sufficiently to be detectable by electrophoresis. A number of protocols have been described for SSCP including, but not limited to, Lee *et al.*, *Anal. Biochem.* 205:289-293 (1992); Suzuki *et al.*, *Anal. Biochem.* 192:82-84 (1991); Lo *et al.*, *Nucleic Acids Research* 20: 1005-1009 (1992); and Sarkar *et al.*, *Genomics* 25: 13:441-443 (1992). It is understood that one or more of the nucleic acids of the present invention, may be utilized as markers or probes to detect polymorphisms by SSCP analysis.

Polymorphisms may also be found using a DNA fingerprinting technique called amplified fragment length polymorphism (AFLP), which is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA to profile that DNA (Vos *et al.*, *Nucleic Acids Res.* 23:4407-4414 5 (1995)). This method allows for the specific co-amplification of high numbers of restriction fragments, which can be visualized by PCR without knowledge of the nucleic acid sequence.

AFLP employs basically three steps. Initially, a sample of genomic DNA is cut with restriction enzymes and oligonucleotide adapters are ligated to the 10 restriction fragments of the DNA. The restriction fragments are then amplified using PCR by using the adapter and restriction sequence as target sites for primer annealing. The selective amplification is achieved by the use of primers that extend into the restriction fragments, amplifying only those fragments in which the primer extensions match the nucleotide flanking the restriction sites. These 15 amplified fragments are then visualized on a denaturing polyacrylamide gel.

AFLP analysis has been performed on *Salix* (Beismann *et al.*, *Mol. Ecol.* 6:989-993 (1997)), *Acinetobacter* (Janssen *et al.*, *Int. J. Syst. Bacteriol.* 47:1179-1187 (1997)), *Aeromonas popoffi* (Huys *et al.*, *Int. J. Syst. Bacteriol.* 47:1165-1171 (1997)), *Phytophthora infestans* (Van der Lee *et al.*, *Fungal Genet. Biol.* 20 21:278-291 (1997)), *Bacillus anthracis* (Keim *et al.*, *J. Bacteriol.* 179:818-824 (1997)), *Astragalus cremnophylax* (Travis *et al.*, *Mol. Ecol.* 5:735-745 (1996)), *Arabidopsis* (Cnops *et al.*, *Mol. Gen. Genet.* 253:32-41 (1996)), *Escherichia coli* (Lin *et al.*, *Nucleic Acids Res.* 24:3649-3650 (1996)), *Aeromonas* (Huys *et al.*, *Int. J. Syst. Bacteriol.* 46:572-580 (1996)), nematode (Folkertsma *et al.*, *Mol. Plant Microbe Interact.* 9:47-54 (1996)), and human (Latorra *et al.*, *PCR Methods Appl.* 25 3:351-358 (1994)). AFLP analysis has also been used for fingerprinting mRNA (Money *et al.*, *Nucleic Acids Res.* 24:2616-2617 (1996), the entirety of which is herein incorporated by reference; Bachem *et al.*, *Plant J.* 9:745-753 (1996)). It is

understood that one or more of the nucleic acids of the present invention, may be utilized as markers or probes to detect polymorphisms by AFLP analysis or for fingerprinting RNA.

The polymorphism obtained by these approaches can then be cloned to 5 identify the mutation at the coding region, which alters the protein's structure or the regulatory region of the gene that affects its expression level. Changes involving promoter interactions with other regulatory proteins can be identified by, for example, gel shift.

In accordance with an embodiment of the invention, a sample DNA is 10 obtained from a patient's cells. In a preferred embodiment, the DNA sample is obtained from the patient's blood. However, any source of DNA may be used. The DNA may be subjected to interrogation to determine the presence or absence of a polymorphism.

15

3. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., antisense, polypeptides, peptidomimetics, small molecules or other drugs) which bind to nucleic acid molecules, polypeptides or proteins described herein or 20 have a stimulatory or inhibitory effect on, for example, expression or activity of the nucleic acid molecules, polypeptides or proteins of the invention.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of protein or polypeptide described herein or biologically active portion thereof. The test compounds of the 25 present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library

method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K. S. (1997) *Anticancer Drug Des.* 5 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.*, 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. U.S.A.*, 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.*, 37:2678; Cho *et al.* (1993) *Science*, 261:1303; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.*, 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.*, 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.*, 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques*, 13:412-421), or on beads (Lam(1991) *Nature*, 354:82-84), 15 chips (Fodor (1993) *Nature*, 364:555-556), bacteria(Ladner U.S. Pat. No. 5,223,409), spores (Ladner U.S. Pat. No. '409), plasmids (Cull *et al.*(1992) *Proc. Natl. Acad. Sci. U.S.A.*, 89:1865-1869) or on phage (Scott and Smith (1990) *Science*, 249:386-390); (Devlin (1990) *Science*, 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.*, 97:6378-6382); (Felici (1991) *J. Mol. Biol.*, 20 222:301-310); (Ladner *supra*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses an encoded protein which is a cell surface receptor is contacted with a test compound and the ability of the test compound to bind to the receptor is determined. The cell, for example, can be of mammalian origin, such as from 25 cartilage tissue or bone. Determining the ability of the test compound to bind to the receptor can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the receptor can be determined by detecting the labeled with .sup.125 I, .sup.35 S,

.sup.14 C, or .sup.3 H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected 5 by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a test compound to interact with the receptor without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a test compound with the receptor without the labeling of either the 10 test compound or the receptor. McConnell, H. M. *et al.* (1992) *Science*, 257:1906-1912. A "microphysiometer" (e.g., Cytosensor.TM.) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between ligand and receptor.

15 In one embodiment, the assay comprises contacting a cell which expresses an encoded protein described herein on the cell surface (e.g., a receptor) with a receptor ligand or biologically-active portion thereof, to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the receptor, wherein determining the ability of 20 the test compound to interact with the receptor comprises determining the ability of the test compound to preferentially bind to the receptor as compared to the ability of the ligand, or a biologically active portion thereof, to bind to the receptor.

25 In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a particular target molecule described herein with a test compound and determining the ability of the test compound to modulate or alter (e.g. stimulate or inhibit) the activity of the target molecule. Determining the ability of the test compound to modulate the activity of the target molecule can be

accomplished, for example, by determining the ability of a known ligand to bind to or interact with the target molecule.

Determining the ability of the known ligand to bind to or interact with the target molecule can be accomplished by one of the methods described above for

5 determining direct binding. In a preferred embodiment, determining the ability of the known ligand to bind to or interact with the target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., intracellular Ca.sup.2+, diacylglycerol, IP.sub.3, etc.), detecting catalytic/enzymatic activity of the target

10 an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a cellular response, for example, development, differentiation or rate of proliferation.

15 In yet another embodiment, an assay of the present invention is a cell-free assay in which protein of the invention or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the protein or biologically active portion thereof is determined. Binding of the test compound to the protein can be determined either directly or indirectly as

20 described above. In one embodiment, the assay includes contacting the protein or biologically active portion thereof with a known compound which binds the protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the protein. Determining the ability of the test compound to interact with the protein

25 comprises determining the ability of the test compound to preferentially bind to the protein or biologically active portion thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which a protein of the invention or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate or alter (e.g., stimulate or inhibit) the activity of the protein or biologically active portion

5 thereof is determined. Determining the ability of the test compound to modulate the activity of the protein can be accomplished, for example, by determining the ability of the protein to bind to a known target molecule by one of the methods described above for determining direct binding. Determining the ability of the protein to bind to a target molecule can also be accomplished using a technology such as real-time Bimolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIACore.TM.). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

10 In an alternative embodiment, determining the ability of the test compound to modulate the activity of a protein of the invention can be accomplished by determining the ability of the protein to further modulate the activity of a target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

15 In yet another embodiment, the cell-free assay involves contacting a protein of the invention or biologically active portion thereof with a known compound which binds the protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the protein, wherein determining the ability of the test compound to interact with the protein comprises determining the ability of the protein to preferentially bind to or modulate the activity of a target molecule.

The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of isolated proteins. In the case of cell-free assays in which a membrane-bound form an isolated protein is used it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the 5 isolated protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton.RTM.X-100, Triton.RTM.X-114, Thesit.RTM., Isotridecyloxy(ethylene glycol ether).sub.n.3-[(3-cholamidopropyl)dimethylammonio]-1-propane sulfonate 10 (CHAPS), 3-[(3-cholamidopropyl)dimethylammonio]2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,Ndimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either the protein or its target 15 molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to the protein, or interaction of the protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such 20 vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtitre plates, 25 which are then combined with the test compound or the test compound and either the non-adsorbed target protein or protein of the invention, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate

wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding or activity determined using standard techniques.

5 Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a protein of the invention or a target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated protein of the invention or target molecules can be prepared from biotin-NHS(N-hydroxysuccinimide) using techniques well known 10 in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with a protein of the invention or target molecules, but which do not interfere with binding of the protein to its target 15 molecule, can be derivatized to the wells of the plate, and unbound target or protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the protein or target molecule.

20 In another embodiment, modulators of expression of nucleic acid molecules of the invention are identified in a method wherein a cell is contacted with a candidate compound and the expression of appropriate mRNA or protein in the cell is determined. The level of expression of appropriate mRNA or protein in the presence of the candidate compound is compared to the level of expression of 25 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of expression based on this comparison. For example, when expression of mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than

in its absence, the candidate compound is identified as a stimulator or enhancer of the mRNA or protein expression. Alternatively, when expression of the mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor 5 of the mRNA or protein expression. The level of mRNA or protein expression in the cells can be determined by methods described herein for detecting mRNA or protein.

In yet another aspect of the invention, the proteins of the invention can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. 10 Pat. No. 5,283,317; Zervos *et al.* (1993) *Cell*, 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.*, 268:12046-12054; Bartel *et al.* (1993) *Biotechniques*, 14:920-924; Iwabuchi *et al.* (1993) *Oncogene*, 8:1693-1696; and Brent W094/10300), to identify other proteins (captured proteins) which bind to or interact with the proteins of the invention and modulate their activity. Such captured proteins are 15 also likely to be involved in the propagation of signals by the proteins of the invention as, for example, downstream elements of a protein-mediated signaling pathway. Alternatively, such captured proteins are likely to be cell-surface molecules associated with non-protein-expressing cells, wherein such captured proteins are involved in signal transduction.

20 The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a protein of the invention is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). 25 In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an protein-dependent

complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be 5 detected, and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the protein of the invention.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to 10 further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a modulating agent, an antisense nucleic acid molecule, a specific antibody, or a protein-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described 15 herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

Animal models appropriate for the study of the progression of OA include 20 the anterior cruciate ligament dog model (Rubin *et al.*, *Clin. Orthop.* 113:212 (1975); Johnson, *Exp. Pathol.* 30:209 (1986); Pelletier *et al.*, *Osteoarthritis Cartilage* 7:416 (1999)), the rabbit partial meniscectomy model (Colombo *et al.*, *Arthritis Rheum.* 26:875 (1983); Butler *et al.*, *Arthritis Rheum.* 26:1380 (1983)), and the guinea pig spontaneous model (Watson *et al.*, *Osteoarthritis Cartilage* 25 4:197 (1996); Jiminez *et al.*, *Lab Anim. Sci.* 47:598 (1997)).

4. Detection Assays

Portions or fragments of the nucleotide sequences identified herein (and the corresponding complete gene sequences) can be used as polynucleotide

reagents. For example, the sequences can be used to map their respective genes on a chromosome and thus locate gene regions associated with genetic disease.

a. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the nucleic acid molecules, described herein, can be used to map the location of the corresponding genes on a chromosome. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, genes can be mapped to chromosomes by preparing PCR primers (preferably 15-30 bp in length) from the nucleic acid molecules described herein. Computer analysis of the sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the appropriate nucleotide sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et*

al. (1983) *Science*, 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the nucleic acid molecules of the invention to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a specified sequence to its chromosome include in situ hybridization (described in Fan, Y *et al.* (1990) *PNAS*, 97:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence in situ hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. for a review of this technique, see Verma *et al.*, *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can

be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal 5 mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, *Medelian Inheritance in Man*, available on-line through Johns Hopkins University Welch 10 Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J *et al.* (1987) *Nature*, 325:783-787.

Moreover, differences in the DNA sequences between individuals affected 15 and unaffected with a disease associated with a specified gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as 20 deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

25 5. Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically.

Accordingly, one aspect of the present invention relates to diagnostic assays for determining protein and/or nucleic acid expression as well as activity of proteins of the invention, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with activity or expression of proteins or nucleic acids of the invention.

For example, mutations in a specified gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with expression or activity of nucleic acid molecules or proteins of the invention.

Another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of proteins of the invention in clinical trials.

These and other agents are described in further detail in the following sections.

20 a. Diagnostic Assays

An exemplary method for detecting the presence or absence of proteins or nucleic acids of the invention in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting the protein, or nucleic acid (e.g., mRNA, genomic DNA) that encodes the protein, such that the presence of the protein or nucleic acid is detected in the biological sample. A preferred agent for detecting mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA sequences described herein. The nucleic

acid probe can be, for example, a full-length nucleic acid, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to appropriate mRNA or genomic DNA. Other suitable probes for use in the 5 diagnostic assays of the invention are described herein.

A preferred agent for detecting proteins of the invention is an antibody capable of binding to the protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab.sup.1).sub.2) can be used. The term 10 "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently 15 labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mRNA, protein, or 20 genomic DNA of the invention in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detection of protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. In vitro techniques for detection 25 of genomic DNA include Southern hybridizations. Furthermore, in vivo techniques for detection of protein include introducing into a subject a labeled anti-protein antibody. For example, the antibody can be labeled with a radioactive

marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules 5 from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample or biopsy isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a 10 compound or agent capable of detecting protein, mRNA, or genomic DNA of the invention, such that the presence of protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of protein, mRNA or genomic DNA in the control sample with the presence of protein, mRNA or genomic DNA in the test sample.

15 The invention also encompasses kits for detecting the presence of proteins or nucleic acid molecules of the invention in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting protein or mRNA in a biological sample; means for determining the amount of in the sample; and means for comparing the amount of in the sample with a standard. 20 The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect protein or nucleic acid.

b. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to 25 identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of proteins and nucleic acid molecules of the invention. For example, the assays described herein, such as the preceding diagnostic assays or the following assays can be utilized to identify a subject

having or at risk of developing a disorder associated with protein or nucleic acid expression or activity such as OA. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing OA. Thus, the present invention provides a method for identifying a disease or disorder associated with 5 aberrant expression or activity of proteins or nucleic acid molecules of the invention, in which a test sample is obtained from a subject and protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the protein or nucleic acid sequence of the invention. As used herein, a "test sample" refers to a 10 biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., synovial fluid), cell sample, or tissue (e.g., cartilage tissue).

Furthermore, the prognostic assays described herein can be used to 15 determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, polypeptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant expression or activity of a protein or nucleic acid molecule of the invention. For example, such methods can be used to determine whether a subject can be 20 effectively treated with an agent for a disorder, such as OA. Alternatively, such methods can be used to determine whether a subject can be effectively treated with an agent for OA. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of a protein or nucleic acid of the present 25 invention, in which a test sample is obtained and protein or nucleic acid expression or activity is detected (e.g., wherein the abundance of particular protein or nucleic acid expression or activity is diagnostic for a subject that can be

administered the agent to treat a disorder associated with aberrant expression or activity.)

The methods of the invention can also be used to detect genetic alterations in genes or nucleic acid molecules of the present invention, thereby determining if

5 a subject with the altered gene is at risk for OA. In one embodiment, the method includes detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a particular protein, or the mis-expression of the gene.

For example, such genetic alterations can be detected by ascertaining the existence

10 of at least one of (1) a deletion of one or more nucleotides; (2) an addition of one or more nucleotides; (3) a substitution of one or more nucleotides, (4) a chromosomal rearrangement; (5) an alteration in the level of a messenger RNA transcript; (6) aberrant modification, such as of the methylation pattern of the genomic DNA; (7) the presence of a non-wild type splicing pattern of a messenger

15 RNA transcript; (8) a non-wild type level; (9) allelic loss; and (10) inappropriate post-translational modification. As described herein, there are a large number of assay techniques known in the art which can be used for detecting alterations in a particular gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

20 In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202), such an anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.* (1988) *Science*, 241:1077-1080; and Nakazawa *et al.* (1994) *PNAS*, 91:360-364), the latter of

25 which can be particularly useful for detecting point mutations (see Abravaya *et al.* (1995) *Nucleic Acids Res.*, 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample

with one or more primers which specifically hybridize to the gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a 5 control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J. C. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:1874-10 1878), transcriptional amplification system (Kwoh, D. Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA*, 86:1173-1177), Q-Beta Replicase (Lizardi, P. M. *et al.*, (1988) *Bio/Technology*, 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the 15 detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a given gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested 20 with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for sample, U.S. Pat. No. 5,498,531) can be used to score for the presence of specific mutations by 25 development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotide probes (Cronin, M. T. *et al.*

(1996) *Human Mutation*, 7:244-255; Kozal, M. J. *et al.* (1996) *Nature Medicine*, 2:753-759). For example, genetic mutations can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M. T. *et al.* *supra*. Briefly, a first hybridization array of probes can be used to scan through 5 long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations 10 detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the gene and detect mutations by 15 comparing the sequence of the gene from the sample with the corresponding wild-type (control) gene sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1997) *PNAS*, 74:560) or Sanger ((1977) *PNAS*, 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic 20 assays ((1995) *Biotechniques*, 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.*, 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.*, 38:147-159).

Other methods for detecting mutations include methods in which 25 protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science*, 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type sequence

with potentially mutant RNA or DNA obtained from a tissue sample. The double-standard duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to base pair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with

5 Rnase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digest the mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example Cotton *et al.* (1988) *Proc. Natl. Acad. Sci. USA*, 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.*, 217:286-295. In a

10 preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair", enzymes) in defined systems for detecting and mapping point mutations in cDNAs obtained from samples of cells. For example,

15 the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis*, 15:1657-1662). According to an exemplary embodiment, a probe based on an nucleotide sequence of the invention is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch

20 repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic

25 mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc. Natl. Acad. Sci. USA*, 86:2766, see also Cotton (1993) *Mutat Res.*, 285:125-144; and Hayashi (1992) *Genet Anal. Tech. Appl.*, 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids will be denatured and allowed to

renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by 5 using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet.*, 7:5).

10 In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature*, 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a 15 GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys. Chem.*, 265:12753).

Examples of other techniques for detecting point mutations include, but are 20 not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature*, 324:163); Saiki *et al.* (1989) *Proc. Natl. Acad. Sci. USA*, 86:6320). Such allele-specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the 25 oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification 5 depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.*, 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech*, 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection 10 (Gasparini *et al.* (1992) *Mol. Cell Probes*, 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci. USA*, 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by 15 looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a 20 disease or illness involving a gene of the present invention. Any cell type or tissue in which the gene is expressed may be utilized in the prognostic assays described herein.

c. Monitoring of Effects During Clinical Trials

25 Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of nucleic acid molecules or proteins of the present invention (e.g., modulation of cellular signal transduction, regulation of gene transcription in a cell involved in development or differentiation, regulation of

cellular proliferation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase gene expression, protein levels, or upregulate protein activity, can be monitored in clinical trials of subjects

5 exhibiting decreased gene expression, protein levels, or downregulated protein activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease gene expression, protein levels, or downregulate protein activity, can be monitored in clinical trials of subjects exhibiting increased gene expression, protein levels, or upregulated protein activity. In such clinical trials,

10 the expression or activity of the specified gene and, preferably, other genes that have been implicated in, for example, a proliferative disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which

15 modulates protein activity (e.g., identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on OA, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of the specified gene and other genes implicated in OA. The levels of gene expression (i.e., a gene expression pattern) can be quantified by

20 Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of the specified gene or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state

25 may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an

agonist, antagonist, peptidomimetic, protein, polypeptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (I) obtaining a pre-administration sample from a subject prior to administration of the agent: (ii) detecting the level of expression of

5 a specified protein, mRNA, or genomic DNA of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the protein, mRNA, or genomic DNA in the pre-

10 administration sample with the protein, mRNA, or genomic DNA in the post-administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of the protein or nucleic acid molecule to higher levels than detected, i.e., to increase effectiveness of the agent.

15 Alternatively, decreased administration of the agent may be desirable to decrease effectiveness of the agent. According to such an embodiment, protein or nucleic acid expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

20

6. Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant expression or activity of proteins or nucleic acids of the invention. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical

development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or

5 therapeutic treatment with the molecules of the present invention or modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug related side effects.

10

a. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with aberrant expression or activity of genes or proteins of the present invention, by administering to the subject an agent which

15 modulates expression or at least one activity of a gene or protein of the invention. Subjects at risk for a disease which is caused or contributed to by aberrant gene expression or protein activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein.

Administration of a prophylactic agent can occur prior to the manifestation of

20 symptoms characteristic of the aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of aberrancy, for example, an agonist or antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the present invention are further

25 discussed in the following subsections.

b. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating expression or activity of genes or proteins of the invention for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of the specified protein associated with the cell. An agent that modulates protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a protein described herein, a polypeptide, a peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more protein activities. Examples of such stimulatory agents include active protein as well as a nucleic acid molecule encoding the protein that has been introduced into the cell. In another embodiment, the agent inhibits one or more protein activities. Examples of such inhibitory agents include antisense nucleic acid molecules and anti-protein antibodies. These modulatory methods can be performed in vitro (e.g., by culturing the cell with the agent) or, alternatively, in vivo (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a protein or nucleic acid molecule of the invention. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) expression or activity of a gene or protein of the invention. In another embodiment, the method involves administering a protein or nucleic acid molecule of the invention as therapy to compensate for reduced or aberrant expression or activity of the protein or nucleic acid molecule.

Stimulation of protein activity is desirable in situations in which the protein is abnormally downregulated and/or in which increased protein activity is likely to have a beneficial effect. Likewise, inhibition of protein activity is

desirable in situations in which the protein is abnormally upregulated and/or in which decreased protein activity is likely to have a beneficial effect.

c. Pharmacogenomics

5 The molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on the protein activity (e.g., gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (e.g., proliferative or developmental disorders) associated with aberrant protein 10 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active 15 drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a molecule of the invention or modulator thereof, as well as tailoring the dosage and/or therapeutic regimen of treatment with such a molecule or modulator.

Pharmacogenomics deals with clinically significant hereditary variations in 20 the response to drugs due to altered drug disposition and abnormal action in affected persons. See e.g., Eichelbaum, M., *Clin Exp Pharmacol. Physiol.*, (1996) 23(10-11):983-985 and Linder, M. W., *Clin. Chem.* (1997) 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic 25 conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency

(G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants). Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1,000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (e.g., a protein or a receptor of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2(NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations.

For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a molecule or modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions

or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a molecule or modulator of the invention, such as a modulator identified by one of the exemplary screening assays described herein.

5 The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

Examples

10 The following examples will illustrate the invention in greater detail, although it will be understood that the invention is not limited to these specific examples. Various other examples will be apparent to the person skilled in the art after reading the present disclosure without departing from the spirit and scope of the invention. It is intended that all such other examples be included within the scope of the appended claims.

Example 1

15 **Identification of Initial ESTs**

Highly selective targets for osteoarthritic tissue were identified from osteoarthritic tissue to a global database of ESTs selected from libraries made from normal human tissue. Three EST libraries of OA cartilage and synovium from 6 different donors (five cartilage donors and one synovium donor) were compared to two EST libraries representing cartilage and synovium from 5 donors (four cartilage donors and one synovium donor) without OA to identify those ESTs which are representative of genes upregulated as a result of OA. The data mining effort used sequence comparison techniques based on BLAST comparison of individual ESTs to evaluate which ESTs were preferentially observed in the target libraries versus control and/or normal libraries. The resulting EST sequences represent osteoarthritis tissue-derived nucleic acids of the invention and can be used to create additional osteoarthritis tissue-derived nucleic acids, proteins,

and polypeptides of the invention. The initial ESTs identified from osteoarthritis tissue libraries (Table 1) were run through bioinformatic computer analysis for the presence of specific structural features, and were given the best possible annotation. Table 1 correlates that information. The sequences which were found 5 to be secreted proteins seen at a frequency which statistically correlates with OA disease presence appear in Table 2.

Table 1. Nucleic acid molecules expressed in osteoarthritis tissue with potential uses for the development of human therapeutics and diagnostics.

10 **Table 2:** Preferred candidate markers of OA Example 2

Example 2

Strategy for obtaining FL clones

The sequences, or fragments of them, disclosed can be used to directly obtain full length clones from cDNA libraries and genomic clones from genomic 15 libraries. A number of methods have been described to obtain cDNA and genomic clones, including 5' RACE (for example, using the Marathon cDNA Amplification Kit from Clonetech, Inc.), Genetrapp (LifeTechnologies, Inc.), colony hybridization (Sambrook *et al.*, 1989. *Molecular Cloning, A Laboratory Manual* 2nd Ed., Cold Spring Harbor Press. Cold Spring Harbor, New York) and 20 array hybridization. One skilled in the art can refer to general reference texts for detailed descriptions of known techniques or equivalent techniques. These texts include *Current Protocols in Molecular Biology* (Ausubel, *et al.*, eds., John Wiley & Sons, N.Y. (1989), and supplements through September 1998), *Molecular Cloning, A Laboratory Manual* (Sambrook *et al.*, 2nd Ed., Cold Spring Harbor 25 Press, Cold Spring Harbor, New York (1989)), and *Current Protocols in Immunology* (Coligan, ed., John Wiley and Sons. Toronto (1994)), for example, each of which are specifically incorporated by reference in their entirety.

A preferred method is array hybridization. A first step in array hybridization is to obtain a high quality library with high complexity and a high proportion of full length or high molecular weight clones. A cDNA library can be purchased from a number of commercial sources or a new library prepared from

5 mRNA derived from tissue known or suspected to express the gene. Details on library construction can be found in Sambrook *et al.* (*Molecular Cloning, A Laboratory Manual* 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York). Similarly, genomic libraries may be purchased from commercial sources.

A plasmid cDNA library constructed from human OA tissue -derived

10 mRNA is plated on agar and a picking robot (for example, a 'Q' BOT from Genetix, Inc.) is used to pick individual colonies into 96 well plates containing LB medium. The isolated *E. coli* cells are grown overnight at 30°C and placed at 4°C until ready for spotting. The *E. coli* are then robotically spotted in high density grids on nitrocellulose membranes overlaying solid agar medium. Preferably,

15 each colony is double spotted at two different locations. The colonies are allowed to grow to 0.1-0.2 cm and then the membranes are prepared for hybridization to nucleic acid sequences. Standard protocols for membrane preparation can be found in Sambrook (*Molecular Cloning, A Laboratory Manual* 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York).

20 Selection of nucleic acid probes, or similarly PCR primers, can be done using a variety of methods, including publicly available programs such as MIT Primer 3 (Rozen and Skaletsky (1996, 1997, 1998)), for example.

Synthetic oligos or a restriction fragment containing probe sequences, from

25 the original clone sequence, are labeled with ^{33}P using conventional means, such as the random hexamer priming method (High Prime kit from Boehringer Mannheim). The labeled probes are then hybridized to the arrayed filters under stringent conditions (Sambrook *et al.*, 1989, *Molecular Cloning, A Laboratory Manual* 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York). In

this case, the wash conditions can approach 70-80°C at 0.1 M NaCl. After washing, the positive colonies can be visualized by exposing the blots to film or to a phosphoimaging screen. The processed image reveals colonies containing cDNA clones of the targeted gene. Alternatively, a non-radioactive method might

5 be chosen to label the probes and detect colony hybridization (for example, the DIG Non-Radioactive DNA Labeling and Detection Kit from Boehringer Mannheim). Each clone is analyzed by restriction digests to identify the longest clone in the group. The longer clones are sequenced using an automated sequencing system (for example a Perkin-Elmer ABI 377) and the sequences

10 evaluated for a complete open reading frame initiated with a methionine start codon (ATG). Genomic clones can also be analyzed for intron-exon boundaries using known methods as well as compared to relevant or homologous gene or protein sequence information to determine the coding regions.

Example 3

15 **Expression and Purification of a Polypeptide in *E. coli* Host**

The vector pProEX HT is used for expression of a polypeptide in a bacterial host system. (LifeTechnologies Inc., Gaithersburg, MD). The plasmid encodes the ampicillin resistance gene ("Apr") and contains a pBR322 origin of replication (ori), an IPTG inducible promoter, a ribosome binding site, and six codons encoding a histidine tag at the amino terminus. The his tag allows affinity purification using immobilized metal ion affinity chromatography (IMAC), such as with the nickel-nitrilo-tri-acetic acid (Ni-NTA) affinity resin. The cloning region contains suitable restriction enzyme cleavage sites for insertion of polypeptide encoding sequences. The vector also encodes a TEV (Tobacco Etch Virus) protease cleavage site to remove the his tag region from the amino terminus of the expressed polypeptide.

The desired polypeptide encoding sequence or fragment, typically lacking any hydrophobic leader sequence, is PCR amplified from a cDNA clone. Primers

designed from at least one of SEQ NO:1-82, which anneal to the amino terminal sequences of the desired polypeptide encoding region, and from the 3' end of the cDNA clone, preferably within or at the beginning of vector sequences of the cDNA clone, are used. Additional sequence containing restriction sites, to 5 facilitate cloning into pProEX HT vector, or stop codons can be incorporated into one or both of the primers.

PCR amplified sequences and the vector are digested with appropriate restriction enzymes and then ligated together. Insertion of the DNA into the digested pProEX HT vector places a polypeptide encoding sequence downstream from the 10 trc promoter and in proper reading frame to an initiator AUG.

The ligation mixture is transformed into competent *E. coli* cells using standard procedures. Sambrook, *et al.*, 1989; Ausubel, 1989, supplements through September 1998. *E. coli* strain DH5a is used, however other strains are possible. Amp resistant colonies indicate successful transformation. Plasmid DNA from 15 resistant colonies is isolated and the correct construction confirmed by restriction analysis, PCR, and DNA sequencing.

Clones containing the desired constructs are grown over night in LB media supplemented with amp (100 mg/ml). These culture are used to inoculate production cultures, where the cells are grown at 370C to an OD590 density of 20 between about 0.4 and 0.6 before induction by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM, and allowed an additional 3 to 4 hours to express polypeptide. Cells then are harvested by centrifugation.

The cell pellet is then brought up in 6M guanidine-HCl, pH8, at 40C, and 25 stirred for 4 hours. The cell debris is removed by centrifugation and the supernatant containing the expressed polypeptide is dialyzed against a refolding buffer, such as a NaCl-based or Tris-based buffer, at about pH 6 and containing protease inhibitors. After refolding, the polypeptide is purified by IMAC and

-81-

optionally cleaved with TEV protease. The cleavage reaction is run through a size exclusion gel to purify the desired polypeptide. Purified polypeptide is stored at 4°C or frozen at -80°C. Gel electrophoresis can be used to verify production of desired polypeptide or for purification..

5

As noted above, the specific examples should not be interpreted as a limitation to the scope of the invention. Instead, they are merely exemplary embodiments one skilled in the art would understand from the entire disclosure of this invention.

10

Table 1*

Seq NO	Hit Description	P Score
SEQ NO 1	Human mRNA for fibronectin (FN precursor	100
SEQ NO 2	Human lumican mRNA, complete cds.	100
SEQ NO 3	Human ferritin H chain mRNA, complete cd	100
SEQ NO 4	Human osteoprotegerin (OPG) mRNA, comple	100
SEQ NO 5	Human DNA for osteopontin, complete cds.	100
SEQ NO 6	Human complement decay-accelerating fact	100
SEQ NO 7	Human SOD-2 gene for manganese superoxid	100
SEQ NO 8	Human chitinase (HUMTC1HIT) mRNA, exon 1b	100
SEQ NO 9	Human CRTI.1 gene exons 1, 2, 3, 4 and 5.	100
SEQ NO 10	Human glia-derived nexin (GDN) mRNA, 5'	100
SEQ NO 11	Human SPARC osteonectin mRNA, complete c	100
SEQ NO12	Human mRNA for CD59, an LY-6-like protei	100
SEQ NO 13	Human mRNA for pro-alpha-1 type 3 collag	100
SEQ NO 14	Human mRNA for eukaryotic initiation fac	100
SEQ NO 15	Human mRNA for DIF-2 protein.	100
SEQ NO 16	Human apolipoprotein J mRNA, complete cd	100
SEQ NO 17	Human chondroitin/dermatan sulfate prote	100
SEQ NO 18	Human alpha-1-antichymotrypsin mRNA, 3'	100
SEQ NO 19	Human alpha-1-antitrypsin mRNA, complete	100
SEQ NO 20	Human mRNA for serine protease.	100
SEQ NO 21	Human mRNA for caveolin.	96
SEQ NO 22	Human G0S2 protein gene, complete cds.	100
SEQ NO 23	Human mRNA for CC chemokine LARC precurs	100
SEQ NO 24	Human mRNA for proliferation-associated	100
SEQ NO 25	Human mRNA for collagen VI alpha-2 C-ter	100
SEQ NO 26	Human fos proto-oncogene (c-fos), comple	100
SEQ NO 27	Human zinc finger transcriptional regula	100
SEQ NO 28	Human zinc finger protein (MAZ) mRNA.	100
SEQ NO 29	Human muscle creatine kinase gene (CKMM)	100
SEQ NO 30	FBLN1: fibulin-1, isoform D precursor	9
SEQ NO 31	Human skeletal muscle LIM-protein FHL1 m	100
SEQ NO 32	Human caldesmon mRNA, complete cds.	100
SEQ NO 33	Human slow skeletal muscle troponin T mR	100
SEQ NO 34	Human mRNA for nuclear envelope protein	100
SEQ NO 35	Human mRNA for titin protein (clone hh1-	100
SEQ NO 36	Human alpha-B-crystallin gene, 5' end.	100
SEQ NO 37	Human suiliso1 mRNA, complete cds.	100
SEQ NO 38	Human CAPL protein mRNA, complete cds.	100
SEQ NO 39	Human mRNA for KIAA0253 gene, partial cd	100
SEQ NO 40	Human tenascin-X (XA) mRNA, complete cds	100
SEQ NO 41	Human gene for fibromodulin.	100
SEQ NO 42	Human RASF-A PLA2 mRNA, complete cds.	100
SEQ NO 43	Human mRNA for activin beta-A, exon1,2.	100

-83-

SEQ NO 44	Human tumor necrosis factor-inducible (T	100
SEQ NO 45	nucleotide pyrophosphohydrolase	100
SEQ NO 46	Human Pig12 (PIG12) mRNA, complete cds.	100
SEQ NO 47	Human mRNA for cysteine dioxygenase, com	100
SEQ NO 48	Human sialoprotein mRNA, complete cds.	100
SEQ NO 49	Human mRNA for alpha-1 type II collagen.	100
SEQ NO 50	Human alpha-2 type XI collagen mRNA (COL	100
SEQ NO 51	F08D12.3	4
SEQ NO 52	Rat chondroadherin mRNA, complete cds.	76
SEQ NO 53	ChM-I; chondromodulin-I	76
SEQ NO 54	mimecan	95
SEQ NO 55	Human chondroadherin gene, 5'flanking re	100
SEQ NO 56	Human bone morphogenetic protein 2A (BMP	30
SEQ NO 57	Human (lambda) DNA for Ig light chain.	17
SEQ NO 58		0
SEQ NO 59		0
SEQ NO 60		0
SEQ NO 61		0
SEQ NO 62		0
SEQ NO 63		0
SEQ NO 64		0
SEQ NO 65		0
SEQ NO 66		0
SEQ NO 67		0
SEQ NO 68		0
SEQ NO 69		0
SEQ NO 70		0
SEQ NO 71		0
SEQ NO 72		0
SEQ NO 73		0
SEQ NO 74		0
SEQ NO 75		0
SEQ NO 76		0
SEQ NO 77	Human mRNA for scrapie responsive protein	78
SEQ NO 78		0
SEQ NO 79		0
SEQ NO 80		0
SEQ NO 81		0
SEQ NO 82		0

-84-

Table 2

Seq NO	Hit Description	P Score
SEQ NO 1	Human mRNA for fibronectin (FN precursor	100
SEQ NO 2	Human lumican mRNA, complete cds.	100
SEQ NO 3	Human ferritin H chain mRNA, complete cd	100
SEQ NO 4	Human osteoprotegerin (OPG) mRNA, comple	100
SEQ NO 5	Human DNA for osteopontin, complete cds.	100
SEQ NO 6	Human complement decay-accelerating fact	100
SEQ NO 8	Human chitinase (HUMTCHIT) mRNA, exon 1b	100
SEQ NO 9	Human CRTL1 gene exons 1, 2, 3, 4 and 5.	100
SEQ NO 10	Human glia-derived nexin (GDN) mRNA, 5'	100
SEQ NO 11	Human SPARC/osteonectin mRNA, complete c	100
SEQ NO 12	Human mRNA for CD59, an LY-6-like protei	100
SEQ NO 30	FBLN1; fibulin-1, isoform D precursor	9

We claim:

1. A nucleic acid having a nucleotide sequence selected from the group consisting of SEQ NO: 1 through SEQ NO: 82, or complements thereof.
- 5 2. The nucleic acid capable of specifically hybridizing to a nucleic acid of claim 1, or a complement thereof.
3. The nucleic acid exhibiting a percentage identity of between about 70% to about 90% with at least a 10 nucleotide region of the sequence of a nucleic acid of claim 2.
- 10 4. The nucleic acid exhibiting a percentage identity of between about 90% to about 99% with at least a 10 nucleotide region of the sequence of a nucleic acid of claim 3.
5. The nucleic acid as claimed in any one of claims 3 or 4 wherein said nucleic acid is detectably labeled.
- 15 6. The nucleic acid of claim 5 wherein said sequence is a marker of osteoarthritis progression.
7. The nucleic acid of claim 5 wherein said label is selected from the group consisting of radioactive, fluorescent, chemi-luminescent, and chromogenic agents, and magnetic particles.
- 20 8. A method of identifying a nucleic acid comprising contacting a hybridization probe as claimed in claim 5 with a sample containing nucleic acid and detecting hybridization to the hybridization probe.
9. A method of identifying a nucleic acid comprising contacting a PCR probe as claimed in claim 5 with a sample containing nucleic acid and producing multiple copies of a nucleic acid that hybridizes to the PCR probe.
- 25

10. A substantially-purified nucleic acid having at least one 10 nucleotide region substantially identical to a sequence identified in Table 1.
11. A recombinant DNA comprising a nucleic acid according to one of claims 1-4, or 8, wherein the recombinant nucleic acid further comprises a promoter or 5 partial promoter region.
12. A host cell containing a nucleic acid as claimed in claim 10.
13. A method for producing and purifying a polypeptide, said method comprising the steps of:
 - 10 a) culturing the host cell of claim 12 under conditions suitable for the expression of the peptide; and
 - b) recovering the polypeptide from the host cell culture.
14. A substantially-purified protein, polypeptide, or fragment thereof, wherein at least one 15 amino acid region is encoded by a nucleic acid as claimed in one of claims 1-4, or 9.
15. 15. An antibody that specifically binds to a purified protein, polypeptide, or fragment thereof, having at least one region of 5 contiguous amino acids encoded by a nucleic acid as claimed in one of claims 1-4 or 10.
16. A transgenic animal having in one or more of its cells an introduced nucleic acid as claimed in one of claims 1-4 or 10, or progeny of the transgenic animal.
20
17. A cell taken from a transgenic animal or its progeny as claimed in claim 16.
18. A composition comprising a nucleic acid as claimed in one of claims 1-3, or a complement thereof.
- 25 19. A method of identifying a biologically active compound or composition comprising contacting the compound or composition with a sample comprising a

protein, polypeptide, or fragment as claimed in 13, and comparing the interaction between the compound or composition and the protein, polypeptide, or fragment with a control.

20. A compound or composition that is detectable in a method of claim 18.
- 5 21. A computer-readable medium having recorded thereon the sequence information of one or more of SEQ NO:1 through SEQ NO: 82 or complements thereof.
- 10 22. A method of identifying a nucleic acid comprising providing a computer-readable medium as claimed in claim 21 and comparing nucleotide sequence information using a computerized means.
23. A substantially-purified nucleic acid molecule which comprises a nucleic acid sequence that is identical to at least 10 nucleotides of a nucleotide sequence selected from the group consisting of SEQ NO: 1 through SEQ NO: 82, or complements thereof.
- 15 24. A substantially-purified nucleic acid molecule which comprises a nucleic acid sequence that is identical to at least 50 nucleotides of a nucleotide sequence selected from the group consisting of SEQ NO: 1 through SEQ NO: 82, or complements thereof.
- 20 25. A substantially-purified nucleic acid molecule which comprises a nucleic acid sequence that is identical to at least 100 nucleotides of a nucleotide sequence selected from the group consisting of SEQ NO: 1 through SEQ NO: 82, or complements thereof.
26. A substantially-purified protein, polypeptide, or fragment thereof, of claim 13 wherein said substantially-purified protein, polypeptide, or fragment thereof is a fusion protein.
27. An antibody of claim 1 that is detectably-labeled.

28. A transformed cell having a nucleic acid molecule of claim 1.
29. A transformed cell having the antisense of a nucleic acid molecule of claim 1.
30. A process for diagnosis or prognosis of osteoarthritis a mammal from the expression of mRNA or cDNA that is identical to at least 20 nucleotides of a nucleotide sequence selected from the group consisting of SEQ NO: 1 through SEQ NO: 82, or complements thereof.
5
31. A method of isolating a nucleic acid that is identical to at least 20 nucleotides of a nucleotide sequence selected from the group consisting of SEQ NO: 1 through SEQ NO: 82, or complements thereof.
10

THIS PAGE BLANK (USPTO)

<110> Searle/Monsanto
Phippard, Deborah
Vasanthakamur, Geetha
Dotson, Stanton
Ma, Xiao-Jun

<120> Osteoarthritis tissue-derived nucleic acids, polypeptides, vectors, and cells

<130> SO-3221 PR

<160> 82

<210> 1

<211> 310

<212> DNA

<213> Homo sapiens

<400> 1

cagaataact ctttctgcac agaccacact gttttgggtc agactcgagg agggaaattcc 60
aatgggtgcct tggccactt ccccttccta tacaacaacc acaattacac tgattgcact 120
tctgagggca gaagagacaa catgaagtgg tggggacca cacagaacta tgatgccgac 180
cagaagtttgcgtt gtttctgccc catggctgccc cacgaggaaa tctgcacaac caatgaaggg 240
gtcatgtacc gcattggaga tcagtggat aagcagcatg acatgggttc acatgtatgag 300
gtgcacgtt 310

<210> 2

<211> 1986

<212> DNA

<213> Homo sapiens

<400> 2

cttgggctgt cctttctccc cacgttccacc tgcaacttcgt tagagagcag tgttcacatg 60
ccacaccaca agatccccac aatgacataa ctccattcag agactggcgt gactggctg 120
ggctccccca ccccccctca gctcttgcata cactcagaat ctggcagcca gttccgtcct 180
gacagagttc acagcatata ttgggtggatt cttgtccata gtgcacatgc tttaagaatt 240
aacgaaagca gtgtcaagac agtaaggatt caaaccatgtt gccaaaaatg agtctaagt 300
catttactct cttcctggca ttgattggat gtaccagtgg ccagtactat gattatgatt 360
ttcccccatac aatttatggg caatcatcac caaactgtgc accagaatgt aactgcccgt 420
aaagctaccc aagtgccatg tactgtgatg agctgaaatt gaaaagtgtt ccaatggcgc 480
ctcctggaat caagtatctt taccttagga ataaccagat tgaccatatt gatgaaaagg 540
cctttgagaa tgtaactgtt ctgcagtggc tcattctaga tcacaacctt ctagaaaaact 600
ccaagataaa agggagagtt ttctctaaat taaaacaact gaagaagctg catataaacc 660
acaacaacct gacagagttt ctggggccac ttcctcaaacc tctggaggat ctgcagctt 720

ctcataacaa gatcacaacaaag ctgggctctt ttgaaggatt ggtaaacctg accttcatcc 780
 atctccagca caatcggtctg aaagaggatg ctgtttcagc tgctttaaa ggtcttaat 840
 cactcgaata ccttgacttg agcttcaatc agatagccag actgccttct gggctccct 900
 gtctctcttc taactctcta cttagacaac aataagatca gcaacatccc tgatgagtt 960
 ttcaagcggtt ttaatgcatt gcagttatctg cgtttatctc acaacgaact ggctgatagt 1020
 ggaataccctg gaaattcttt caatgtgtca tccctgggtt agctggatct gtcctataac 1080
 aagcttaaaa acataccaaac tgtcaatgaa aaccttggaaa actattacct ggaggtcaat 1140
 caacttgaga agtttgacat aaagagcttc tgcaagatcc tggggccatt atcctactcc 1200
 aagatcaagc atttgcgttt ggatggcaat cgcatttcag aaaccagtct tccaccggat 1260
 atgtatgaat gtctacgtgt tgcttaacgaa gtcactctta attaataatct gtatcctgga 1320
 acaatatttt atggttatgt ttttctgtgt gtcagtttc atagttatcca tatttttatta 1380
 ctgtttatta cttccatgaa ttttaaaatc tgagggaaat gttttgtaaa catttttattt 1440
 ttttaaagaa aagatgaaag gcaggcctat ttcatcacaa gaacacacac atatacacga 1500
 atagacatca aactcaatgc tttatttgta aatttagtgt ttttttattt ctactgtcaa 1560
 atgatgtgca aaacctttta ctgggttgcattt gggaaatcagc caagtttat aatccttaaa 1620
 tcttaatgtt cctcaaagct tggattaaat acatatggat gttactctct tgacccaaat 1680
 tatcttgata cattcaaattt tgcgttgcattt aaaaataggt ggttagatatt gaggccaaga 1740
 atattgcaaa atacatgaag cttcatgcac ttaaagaagt attttttagaa taagaatttg 1800
 cataacttacc tagtgaaact tttctagaat tatttttac tctaaatgtcat gtatgtttct 1860
 ctttgattat ttgcattgttta tgtttaataa gctacttagca aaataaaaca tagcaaatgg 1920
 catcactgtg tttgacttct tgcgttgcattt ctactgttgcattt tataaaaaat acataaaaca 1980
 atagat 1986

<210> 3
 <211> 920
 <212> DNA
 <213> Homo sapiens

<400> 3

ccgagagtcg tcggggtttc ctgtttcaac agtgcttggaa cggaaacccgg cgctcggttcc 60
 ccaccccccgc cggccgcccc tagccagcccc tccgtcacct cttcacccgca ccctcggtact 120
 gcccccaaggc cccccccgcgc gctccagcgc cgcgcagcca cccggccgc cggccctct 180
 ccttagtcgc cgcctatgacg accgcgttcca cctcgcaggt ggcgcagaac taccaccagg 240
 actcagagggc cgcctatcaac cgcctatgacg acctggagct ctacgcctcc tacgtttacc 300
 tgcgttgcattt tgcgttgcattt gacccgcgttgcgttgcattt gacccgcgttgcattt gacccgcgttgcattt 360

ttcttcacca atctcatgag gagagggAAC atgctgagaa actgatgaag ctgcagaacc 420
 aacgaggtgg ccgaatcttc cttcaggata tcaagaaACC agactgtgat gactgggaga 480
 gcgggctgaa tgcaatggag tgtgcattac atttggaaaa aaatgtaat cagtcactac 540
 tggaaactgca caaactggcc actgacaaaa atgacccca tttgtgtgac ttcattgaga 600
 cacattacct gaatgagcag gtgaaAGCCA tcaaagaATT gggtgaccac gtgaccaact 660
 tgcgcaagat gggagcgcCCC gaatctggct tggcggata tctctttgac aagcacacCC 720
 tgggagacag tgataatgaa agctaaggCT cgggctaATT tccccatAGC cgtggggTGA 780
 ctccctggT caccaaggca gtgcattGcat gttggggTTT ccttacCTT ttctataAGT 840
 tgtaccaaaaa catccactta agttcttga tttgtaccat tccttcaaAT aaagaaATT 900
 ggtacccagg aaaaaaaaaaa 920

<210> 4
 <211> 2139
 <212> DNA
 <213> Homo sapiens

<400> 4

caggcgatac ttccTgttgc cgggacgcta tatataacgt gatgagcgca cgggctgcgg 60
 agacgcacccg gagcgctcgc ccagccgccc cctccaaAGCC cctgaggTTT cggggacca 120
 caatgaacaa cttgctgtgc tgcgcgcTTc gtgtttctgg acatctccat taagtggacc 180
 acccaggaaaa cgtttccTCC aaagtacCTT cattatgacg aagaaacCTC tcattcagCTG 240
 ttgtgtgaca aatgtccTCC tggTaccTAC ctAAAACAAC actgtacAGC aaagtggaaAG 300
 accgtgtgcg ccccttgcCC tgaccactac tacacagaca gctggcacac cagtgacgag 360
 tgtctataCT gcagccccgt gtgcaggAG ctgcagtacg tcaaggcAGGA gtgcatacgc 420
 acccacaacc gcgtgtgcga atgcaaggAA gggcgctacc ttgagataga gttctgcttg 480
 aaacatagga gctgcccTCC tggatttggA gtggTgcAAG ctggAACCC agagcgaaAT 540
 acagTTTgca aaagatgtcc agatgggTTC ttctcaaATg agacgtcatc taaagcacCC 600
 tgttagaaaAC acacaaATTG cagtgtCTT ggtctcCTgc taactcagAA aggAAatgca 660
 acacacgaca acatATgtTC cggAAacAGT gaatcaactc AAAAATgtgg aatagatgtt 720
 accctgtgtg aggaggcatt cttcaggTTT gctgttccTA caaAGTTAC gcctaactgg 780
 cttagtgTCT tggtagacAA ttgcctggc accAAAGTAA acgcagAGAG tggtagAGG 840
 ataaaacggc aacacagCTC acaAGAAACAG acttccAGC tgctgaAGTt atggAAACAT 900
 caaaacAAAG accaAGATAT agtcaAGAAAG atcatccAAg atattgacCT ctgtgAAAAC 960
 agcgtgcAGC ggcacattgg acatgctaAC ctcacCTCG agcagCTCG tagCTTgATG 1020
 gaaagCTTAC cgggAAAGAA agtgggAGCA gaagacATTg aaaaaacaAT aaaggcatgc 1080

aaaccaggc accagatcct gaagctgctc agtttggc gaataaaaaa tggcgaccaa 1140
gacaccccttga agggccta at gcacgcacta aagcactgca aagacgtacc actttcccaa 1200
aactgtca ct cagagtctaa agaagaccat caggccctt cacagcttca caatgtacaa 1260
attgtatcg aagttat ttt tagaaatgtat aggttaaccag gtccaatcg taaaaataag 1320
ctgcttataa ctggaaatgg ccattgagct gttccctcac aattggcgag atcccatgga 1380
tgagtaaact gtttctcagg cacttgaggc tttcagtgtat atctttctca ttaccagtga 1440
ctaatttgc cacagggtac taaaagaaac tatgtatgtgg agaaaggact aacatctcct 1500
ccaataaacc ccaaatgggtt aatccaaactg tcagatctgg atcgatatct actgactata 1560
ttttccctta ttactgctt cagtaattca actggaaatt aaaaaaaaaa aactagactc 1620
cattgtgcct tactaaatat gggaaatgtct aacttaaataa gctttgagat ttca gctatg 1680
ctagaggcctt ttattagaaa gccatatttt tttctgtaaa agttactaat atatctgtaa 1740
cactattaca gtattgctat ttatattcat tcagatataa gatttgatacattatcatc 1800
ctataaagaa acggtatgac ttaatttttag aaagaaaatt atattctgtt tattatgaca 1860
aatgaaaagag aaaatataa ttttaatgg aaagtttgc gcattttctt aataggtact 1920
gccatatttt tctgtgtgg a gtattttat aattttatct gtataagctg taatatcatt 1980
ttatagaaaa tgcattattt agtcaattgt ttaatgttgg aaaacatatg aaatataaaat 2040
tatctgaata ttagatgctc tgagaaaattg aatgtacctt atttaaaaaga ttttatggtt 2100
ttataactat ataaatgaca ttattaaagt ttcaattt 2139

<210> 5
<211> 157
<212> DNA
<213> Homo sapiens

<400> 5

cccaatacta agctccctcg gtttagagcca gccatgagag aaactccaaag tacttctgac 60
tggttctctc tctactcatc cacccttag gtggctgcag aaggaactct gtgcaacccc 120
cagagttctc attctcagtgc acagggaaat gtaatga 157

<210> 6
<211> 2263
<212> DNA
<213> Homo sapiens

<223> unsure at all n locations
<400> 6

acctctgacc acaacaaacc cctactccac ccggcttgc ttgtcccacc cttgggtgacg 60
cagagccccca gcccagaccc cgcccaaagc actcattaa ctggattgc ggancacgag 120

5

gcttctgctt actgcaactc gctccggccg ctgggcgtag tgcgactcgg cggagtcgg 180
gcggcgcgtc cttgttctaa cccggcgcgc catgaccgtc gcgcggccga gcgtgcccgc 240
ggcgctgccc ctccctcgggg agctgccccg gctgctgctg ctggtgctgt tgtgcctgcc 300
ggccgtgtgg ggtgactgtg gccttccccc agatgtaccc aatgcccagc cagctttgga 360
aggccgtaca agttttcccg aggatactgt aataacgtac aaatgtgaag aaagctttgt 420
gaaaatttccct ggcgagaagg actcagtgtat ctgccttaag ggcagtcaat ggtcagatat 480
tgaagagttc tgcaatcgta gctgcgaggt gccaacaagg ctaaattctg catccctcaa 540
acagccttat atcactcaga attatttcc agtcggtaact gttgtggaat atgagtgcgg 600
tccaggttac agaagagaac cttctctatc accaaaaacta acttgccttc agaatttaaa 660
atggtccaca gcagtcgaat ttgtaaaaa gaaatcatgc cctaattccgg gagaaatacg 720
aaatggtcag attgatgtac caggtggcat attatttgtt gcaaccatgc tccttctcat 780
gtaacacagg gtacaaatttta ttggctcga cttcttagttt ttgtcttatt tcaggcagct 840
ctgtccagtg gagtgacccg ttgccagagt gcagagaaat ttattgtcca gcaccaccac 900
aaattgacaa tggaaaatttta caaggggaac gtgaccatata tggatatacg cagtcgtaa 960
cgtatgcatttgc taataaagga ttcaccatga ttggagagca ctctatttttgtactgtga 1020
ataatgtatgc aggagagtgg agtggccac cacatgtatgc cagagggaaaa tctctaactt 1080
ccaaggtccc accaacagtt cagaaaccta ccacagtaaa tggccaaactt acagaagtct 1140
caccaacttc tcagaaaaacc accacaaaaa ccaccacacc aaatgtctaa gcaacacgg 1200
gtacacccgtt ttccaggaca accaaggcatt ttcattgttccaaac aaccccaat aaaggaagtgc 1260
gaaccacttc aggtactacc cgtcttctat ctgggcacac gtgtttcagc ttgacaggtt 1320
tgcttggac gctagtaacc atgggcttgc tgacttagcc aaagaagagt taagaagaaa 1380
atacacacaa gtatacagac tgttcttagt ttcttagact tatctgcata ttggataaaaa 1440
taaatgcaat tgcgtcttc atttaggatg ctttcattgt cttaagatg tgtaggaat 1500
gtcaacagag caaggagaaa aaaggcagtc ctggaaatcac attcttagca cacctacacc 1560
tcttgaaaat agaacaactt gcagaattga gagtgattcc ttccctaaaa gtgtaaagaaa 1620
gcatagagat ttgttcgtat tttagatggg atcacgagga aaagagaagg aaagtgtattt 1680
ttttccacaa gatctgtat gttatccat cttataaagg aaataaaaaa tgaaaaacat 1740
tatttgata tcaaaagcaa ataaaaacc aattcagtct ctctcaagca aaattgtctaa 1800
agagagatga accacattat aaagtaatct ttggctgtaa ggcattttca tcttccttc 1860
gggttggcaa aatattttaa aggtaaaaca tgctggtaa ccaggggtgt tgatgggtat 1920
aaggggagggaa tatagaatga aagactgaat cttcccttgc tgcacaaata gagttggaa 1980
aaagcctgtg aaaggtgtct tctttgactt aatgtcttta aaagtatccca gagatactac 2040

6

aatattaaca taagaaaaga ttatataatta tttctgaatc gagatgtcca tagtcaaatt 2100
 tggaaatctt attctttgt aatatttatt tatatttatt tatgacagtg aacattctga 2160
 ttttacatgt aaaacaagaa aagttgaaga agatatgtga agaaaaatgt atttttccta 2220
 aatagaaata aatgatccca tttttggta aaaaaaaaaa aaa 2263

<210> 7
 <211> 712
 <212> DNA
 <213> Homo sapiens

<400> 7

cttaaaccta ttttagtaatg ttttcccaag tttttttt attttaatt ttttcccaa 60
 gtttattttt ctatttttt ttcattggaaa aatggggtaa ctttagcagtt tcaatattga 120
 agactgaagt ttaaaaaaaaaa tttaaattca aggtactttt aaaattcagt tagaaaaagta 180
 ggctttaaaa attatttagag acaagagtagc caaagcggtg tttgtatgtg tttgtgtgt 240
 tgcattgttg tggattggaa aaacttttggaa gactgattac ttttcaattat atatgtgtca 300
 cagtgaaaca gcttttatgt gtcattgtaa attattgtt gctctctaa ggaaggcgt 360
 gactgtttaa atagacgggc aaggtgaaac cttttgaaag atgagctttt gaatataagt 420
 tttttttttt atcatggttt gtattgaact aacaagggtt gcagatctgc tttttttat 480
 aaagctttttt gattcctact aagctttaag attttttttt ttttcaatgt ttttcaatgt 540
 gtggggctct atttttgtt tggctttctg gttttttttt gttttttttt gttttttttt 600
 actaaagtttgc tttttttttt tttttttttt tttttttttt aagagactaa gggattttac 660
 aaactaaaga ttttagtcat ctggggaaa aggagactttt aagattgtttt ag 712

<210> 8
 <211> 1474
 <212> DNA
 <213> Homo sapiens

<400> 8

ctcaggataaaagaccta gagaatgtgtt tttttttttt aagctggccca aggatatggg 60
 agcaaccaccatgggaccag aagtctctttt gggcaggtg tagtggctttt gttttttttt 120
 cagggaggga tctgcctaca aactggtttgc tttttttttt aactgggtcc caggaccggc 180
 aggaaccagg aaaattcacc cctgaggaat attgaccctt tttttttttt tttttttttt 240
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 300
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 360
 ccattggagg gtacctgtttt gttttttttt tttttttttt tttttttttt tttttttttt 420
 cacgcttggaa attcattaac tccataatcc tttttttttt tttttttttt tttttttttt 480

7
tggatgttaag ctggatctac ccagatcaga aagaaaaacac tcatttcact gtgctgattc 540
atgagtttagc agaaggcttt cagaaggact tcacaaaatc caccaaggaa aggcttctct 600
tgactgcggg gggtatctgc agggaggcaa atgattgata acagctatca agttgagaaaa 660
ctggcaaaag atctggattt catcaacctc ctgtcctttg acttccatgg gtcttggaa 720
aagcccccta tcactggcca caacagccct gctgagcaag gggtggcagg acagagggcc 780
aagctcctac tacaatgtgg aatatgctgt ggggtactgg atacataagg gaatgccatc 840
agagaaggtg gtcatggca tccccacata tggggcactc cttcacactg gcctctgcag 900
aaaccaccgt gggggccccc gcctctggcc ctggagctgc tggacccatc acagagtctt 960
caggcttccct ggcttattat gagatctgcc agttcctgaa aggagccaag atcacgcggc 1020
tccaggatca gcaggttccc tacgcagtca aggggaacca gtgggtggc tatgatgatg 1080
tgaagagtat ggagaccaag gttcagttct taaagaattt aaacctggga ggagccatga 1140
tctggtctat tgacatggat gacttcactg gcaaattctg caaccaggc ctttaccctc 1200
ttgtccaagc agtcaagaga agccttgct ccctgtgaag gattaactta cagagaagca 1260
ggcaagatga ctttgctgcc tggggcctgc tctctccctg gaattctcat gtgggattcc 1320
ccttgcagg ccggcctttg gatctctttt ccaaggcttt cctgacttcc tcttagatca 1380
tagattggac ctggtttgt tttcctgcag ctgttgactt gttggccctga agtacaataa 1440
aaaaaattca ttttgctcca gtaaaaaaaaaaaa aaaa 1474

<210>	9
<211>	592
<212>	DNA
<213>	<i>Homo sapiens</i>

<223> unsure at all n locations
<400> 9

<210> 10
<211> 2004
<212> DNA
<213> Homo sapiens

<400> 10

gcgaccgccc cctgtgatcc agcgagcgcg gtcgtccttg gtggaaggaa ccatgaactg 60
gcatctcccc ctcttcctct tggcctctgt gacgctgcct tccatctgct cccacttcaa 120
tcctctgtct ctcgaggaac taggctccaa cacggggatc caggtttca atcagattgt 180
gaagtcgagg cctcatgaca acatcgat ctctcccat gggattgcgt cggcctggg 240
gatgcttcag ctgggggcgg acggcaggac caagaagcag ctcgccatgg tgatgagata 300
cggcgtaaat ggagttggta aaatattaaa gaagatcaac aaggccatcg tctccaagaa 360
gaataaaagac attgtgacag tggctaacgc cgtgtttgtt aagaatgcct ctgaaattga 420
agtgcctttt gttacaagga acaaagatgt gttccagtgt gaggtccgga atgtgaactt 480
tgaggatccc agcctctgcc tgtgattcca tcaatgcatg gttaaaaaac gaaaccaggg 540
atatgattga caatctgctg tccccagatc ttattgatgg tgtgctcacc agactggtcc 600
tcgtcaacgc agtgttatttc aagggtctgt gaaaatcagc gttccaaaccc gagaacacaa 660
agaaacgcac ttctgtggca gccgacggga aatcctatca agtgc当地 ctggcccage 720
tctccgtgtt cgggtgtggg tcgacaagtg cccccatga tttatggtag aacttcattg 780
aactgcccata ccacggggaa agcatcagca tgctgattgc actgcccact gagagctcca 840
ctccgctgtc tgccatcatc ccacacatca gcaccaagac catagacagc tggatgagca 900
tcatggtgcc caagagggtg caggtgatcc tgcccaagtt cacagctgta gcacaaacag 960
atttgaagga gccgctgaaa gttcttggca ttactgacat gttttagtca tcaaaggcaa 1020
attttgcaaa aataacaagg tcagaaaacc tccatgtttc tcataatcttgc caaaaagcaa 1080
aaattgaagt cagtgaagat ggaaccaaag cttagcagc aacaactgca atttcattg 1140
caagatcatac gcctccctgg tttatagtag acagaccttt tctgttttc atccgacata 1200
atccctacagg tgctgtgtta ttcatggggc agataaacaacc accctgaaga gtatacaaaa 1260
gaaaccatgc aaagcaacga ctactttgct acgaagaaag actcctttcc tgcatcttgc 1320
atagttctgt taaatatttt tgtacatgc ttcttttca aaacttagttc ttaggaacag 1380
actcgatgca agtgtttctg ttctggggagg tattggaggg aaaaaacaag caggatggct 1440
ggaacactgt actgaggaat gaatagaag gcttccagat gtctaaaaga ttctttaaac 1500
tactgaactg ttaccttaggt taacaaccct gttgagttt tgctgtttgt ccagttcagg 1560

aattttgtt ttgtttgtc tatatgtgcg gctttcaga agaaatttaa tcagtgtgac 1620
agaaaaaaaaa atgttttatg gtagctttta cttttatga aaaaaaaaaattt atttgccttt 1680

9
taaattcttt tcccccatcc ccctccaaag tcttgcatacg aagcgttatt ttgggggtag 1740
aaacggtgaa atctctagcc tctttgtgtt tttgttgtt tggttgtgt tgttttat 1800
aatgcataatgtt ttcactaaaa taaaattaa aaaactcctg tcttgcatacg caagggttgc 1860
gttgcgtcgtt gtgcgtgtca ctactggtct gtactcctt gatggcatt ttgttatttt 1920
gtacaaagta aaaataaaact gttatgatgtt gtaaaaataa agctatttct ctgttattt 1980
aaaataaaaaa aaaaaaaaaa aaaa 2004

<210> 11
<211> 2128
<212> DNA
<213> Homo sapiens

<400> 11

agactgcccgg agagcgcgct ctgcgtccgg cctgcctgcc tgccactgag ggttcccagg 60
accatgaggg cctggatctt ctttctcctt tgccgtggccgg ggagggccctt ggcagccccc 120
cagcaagaag ccctgcctga tgagacagag gtgggtggaaag aaactgtggc agaggtgact 180
gaggtatctg tgggagctaa tctgtccag gtggaaagtag gagaattttaa tgatggtgca 240
gaggaaaccg aagaggaggt ggtggcgaa aatccctgcc agaaccacca ctgcaaacac 300
ggcaagggtgt gcgagctgga tgagaacaac acccccatgt gcgtgtgccca ggacccacc 360
agctgcccag ccccccattgg cgagttttag aaggtgtgca gcaatgacaa caagaccc 420
gactcttcctt gccacttctt tgccacaaaag tgccacccctgg agggcaccaaa gaagggccac 480
aagctccacc tggactacat cgggccttgc aaatacatcc ccccttgccct ggactctgag 540
ctgaccgaat tccccctgctg catgcgggac tggctcaaga acgtccctgtt caccctgtat 600
gagagggatg aggacaacaa ctttctgact gagaagcaga agctgcgggt gaagaagatc 660
catgagaatg agaagcgccct ggaggcagg agaccacccc gtggagctgc tggcccccgg 720
cttcgagaag aactataaca tgtacatctt ccctgtacac tggcagttcg gccagctgga 780
ccagcacccccc attgacgggtt acctctccca caccgagctg gtcactgc gtgtccccc 840
catccccatg gacatttgca ccaccccgctt tttcgagacc tgtgacctgg acaatgacaa 900
gtacatcgcc ctggatgagt gggccggctg cttcgccatc aagcagaagg atatcgacaa 960
ggatcttgcgtt atctaaatcc actccttcca cagttccggaa ttctctcttt aaccctcccc 1020
ttcgtgtttc ccccaatgtt taaaatgttt ggtatggttt tggttctgcc tggagacaaag 1080
gtgcttaacat agatTTAAGT gaatacatta acgggtctaa aaatgaaaat tctaacccaa 1140
gacatgacat tcttagctgt aacttaacta ttaaggccctt ttccacacgc attaatagtc 1200
ccatTTTCTT cttggccatTTT gtagcttgc ccattgtctt attggcacat ggggtggacac 1260
ggatctgctg ggatctgcct taaacacaca ttgcagcttc aactttctc ttttagtgttc 1320

10

tgtttcaaact taataacttac cgagtcagac tttgtgttca tttcatttca gggctttggc 1380
 tgcctgtggg ctttccccag ggtggcctgg gaggtgggca aaggaaagta acagacacac 1440
 gatgttgcata aggtatggttt tgggactaga ggctcagtgg tgggagagat ccctgcagaa 1500
 cccaccaacc agaacgtggt ttgcctgagg ctgttaactga gagaaagatt ctggggctgt 1560
 cttatgaaaa tatagacatt ctcacataag cccagttcat caccattcc tcctttacct 1620
 ttcagtgcag tttctttca cattaggctg ttggttcaaa cttttggag cacggactgt 1680
 cagttctctg ggaagtggc agcgcacccct gcagggcttc tcctcctctg tctttggag 1740
 aaccaggcct cttctcaggc gctctaggga ctgccaggct gttcagcca ggaaggccaa 1800
 aatcaagagt gagatgtaga aagttgtaaa atagaaaaag tggagttggt gaatcggttg 1860
 ttcttcctc acatttggat gattgtcata aggttttag catgttcctc cttttcttca 1920
 ccctccccctt tggtcttcta ttaatcaaga gaaacttcaa agttaatggg atggtcggat 1980
 otcacaggct gagaactcgt tcaccccaa gcatttcatg aaaaagctgc ttcttattaa 2040
 tcatacaaacc tctcaccatg atgtgaagag tttcacaaat ctttcaaaat aaaaagtaat 2100
 gacttagaaaa ctgcaaaaaa aaaaaaaaaa 2128

<210> 12
 <211> 2073
 <212> DNA
 <213> Homo sapiens

<400> 12

agtacacact ggggcttata gggactgagc ctactcaagg gtatgtggc ctgtgggtca 60
 gagctggggc atggcaggcg attcagtgtg ctttgcactcc ccctgttaat gttcctctca 120
 gaagccttct tggccttcca gcccctgggtt tttgagacaa ccagcagtca tttgttcgtt 180
 cctgacatcc ctccctgtcc ctcccttcca gtttctgtgg acaatcacaa tggaaatcca 240
 aggagggctc gtccctgttcg ggctgctgct cgtccctggct gtcttctgcc attcaggtca 300
 tagcctgcag tgctacaact gtccctaaacc aactgctgac tgcaaaacag ccgtcaattg 360
 ttcatctgtat ttgtatgcgt gtctcattac caaagctggg ttacaagtgt ataacaagtg 420
 ttggaaagttt gaggcattgca atttcaacga cgtcacaacc ccgcttgagg gaaaatgagc 480
 taacgtacta ctgctgcaag aaggacctgt gtaactttaa cgaacagctt gaaaatggtg 540
 ggacatcctt atcagagaaaa acagttctc tgctggtgac tccatttctg gcagcagcc 600
 ggagccttca tccctaaagtc aacaccagga gagcttctcc caaactcccc gttcctgct 660
 agtcccgcttt ctcttgcgtc cacattctaa aggcttgata ttttccaaat ggatcctgtt 720
 gggaaagaat aaaattagct tgagcaaccc ggctaaagata gaggggctct gggagacttt 780
 gaagaccagt cctgtttgca gggaaagcccc acttgaagga agaagtctaa gagtgaagta 840

11

ggtgtgactt gaactagatt gcatgcttcc tcctttgctc ttggaaagac cagtttgc 900
 agtgcacgct tgagtgggtt ctctgcagcc ctcagattat ttttctctg gtccttgga 960
 ttagtcagt tagcatcatt agtacatctt tggaggggtgg ggcaggagta tatgagcattc 1020
 ctctctcaca tggAACGCTT tcataaactt cagggatccc gtgttgcatt ggaggcatgc 1080
 caaatgttcc atatgtgggt gtcagtcagg gacaacaaga tccttaatgc agagctagag 1140
 gacttctggc agggaaagtgg ggaagtgttc cagatagcag ggcattaaaa cttagagagg 1200
 tacaagtggc tgaaaatcga gttttcctc tgcattttaaa ttttatatgg gctttgttat 1260
 cttccactgg aaaaagtgtaa tagcatacat caatggtgtg ttaaagctat ttccttcct 1320
 ttttttatt ggaatggtag gatatcttgg ctttgcaca cacagttaca gagtgaacac 1380
 tctactacat gtgactggca gtatataatg tgcttatttt aatgttact ggttagaaagg 1440
 cagttcaggt atgtgtgtat atagttgaa tgcaatgggg acacccttg tggttacagt 1500
 ttgagacttc caaaggcatc ccttaataac aacagatctg caggggtatg ttttaccatc 1560
 tgcattccagc ctccctgctaa ctcctagctg actcagcata gattgtataa aataccttg 1620
 taacggctct tagcacactc acagatttt gaggcttca gaagcttcc taaaaatga 1680
 tacacacctt tcacaagggc aaacttttc cttttccctg tgcattttttt tgaatgaatc 1740
 tcaagattca gtagacctaa tgacattgtt attttatgtat cttggctgtt ttaatggca 1800
 taggctgact ttgcagatg gagaaatttc ttgatataatg ttgaaaaaaaa acccttgatt 1860
 atactctgtt ggacaaaccg agtgcattttt atgtatgtt tctgaaaatg aaatataaca 1920
 agtgggtgaa tgcgtttatg gcccggaaagg atatgcagta tgcttaatgg tagcaactga 1980
 aagaagacat cctgagcagt gcccggatc ttctgttgc gcccgttccctt gaacatagga 2040
 aaatagaaaac ttgcattatca aaactttaaaa aaa 2073

<210> 13
 <211> 253
 <212> DNA
 <213> Homo sapiens

<400> 13

gctggctact ttcgcctctg cttcatccca ctattatttt ggcacaacag gaagctgtt 60
 aaggaggatg ttcccatctt ggtcagtcct atgcggatag agatgtctgg aagccagaac 120
 catgccaaat atgtgtctgt gactcaggat ccgttctctg cgatgacata atatgtgac 180
 atcaagaatt agactgcccc aacccagaaaa ttccattttgg agaatgttgc gcagtttgc 240
 cacagcctcc aag 253

<210> 14
 <211> 1749
 <212> DNA

12

<213> Homo sapiens

<223> unsure at all n locations
<400> 14

tcatgtctgc gagccaggat tcccgatcca gagacaaatgg ccccgatggg atggagcccg 60
 aaggcggtcca tcgagagtaa ctggaatgag attgttgaca gctttgatga catgaacctc 120
 tcggagtcgg ttctccgtgg catctacgcc tatggttttg agaagccctc tgccatccag 180
 cagcgagcca ttctaccttg tatcaagggt tatgatgtga ttgctcaagc ccaatctggg 240
 actgggaaaa cggccacatt tgccatatacg attctgcagc agattgaatt agatctaaaa 300
 gccacccagg ccttggtcct agcaccact cgagaattgg ctccagcagat acagaaggcg 360
 gtcatggcac taggagacta catgggcgcc ttctgtcacg cctgtatcg gggcaccaac 420
 gtgcgtgctg aggtgcagaa actgcagatg gaagctcccc acatcatcgt gggtacccct 480
 ggcgcgtgtt ttgatatgct taaccggaga tacctgtccc ccaaatacat caagatgttt 540
 gtactggatg aagctgacga aatgttaagc ctggatcca aggaccagat ctatgacata 600
 ttccaaaagc tcaacagcaa cacccaggtt gttttgtgt cagccacaat gccttctgat 660
 gtgcgtgagg tgaccaagaa gttcatgagg gaccccatcc ggattcttgc caagaaggaa 720
 gagttgaccc tggagggtat ccgccatgtt tacatcaacg tggaaacgaga ggagtggaaag 780
 ctggacacac tatgtgactt gtatgaaacc ctggatcca cccaggcagt catcttcata 840
 aacacccgga ggaagggtgga ctggctcacc gagaagatgc atgctcgaga ttctactgta 900
 tccgcatgc atggagatat ggacaaaag gaacgagacg tgattatgag ggagtttcgt 960
 tctggctcta gcagagttt gattaccact gatccatgg ccagaggcat tgatgtgcag 1020
 caggtttctt tagtcatcaa ctatgaccc cccaccaaca gggaaaacta tatccacaga 1080
 atcggtcgag gtggacgggt tggccgtaaa ggtgtggcta ttaacatggt gacagaagaa 1140
 gacaagagga ctcttcgaga cattgagacc ttctacaaca cctccattga ggaaatgccc 1200
 ctcaatgttgc tgcacccatc ctgaggggtt gatccatgg ccagccccag ccaggcgtca 1260
 atctctgggg gctgaggagc agcaggagg gggaggaaag ggagccaaagg gatggacatc 1320
 ttgtcatttt ttttcttgc ataaatgtca tttttgcagg caaaagaagg aaccgtgaac 1380
 attttagaca cccttttctt tggggtaggc ttgtccca ggcgcncggct cttctccnaa 1440
 aaaaaaaaaa cactaatcca ttccctaaac ctatgaaacct ccagatccca gaggctctcc 1500
 tcacccatgc tgagtcctt tgaaagtgtat tcaaggact atgtcactca gcctcatttg 1560
 ctggacccaa tctggaggga gaacccctaa aaccctaaag tgaggttgcc cagggggttt 1620
 tccccaggtt gggggaaagca ggggagagaa aatggtagcc attttacat tgtttgc 1680
 agtatttttattt gattcaggaa acaaacacaa aatctgaat aaaatgactt ggaaactgaa 1740
 aaaaaaaaaa

1749

<210> 15
 <211> 1232
 <212> DNA
 <213> Homo sapiens

<400> 15

ttacactccg ctcggctcac catgtgtcac tctcgagct gccacccgac catgaccatc 60
 ctgcaggccc cgaccccccgc cccctccacc atccccggac cccggcggggg ctccggtcct 120
 gagatcttca ccttcgaccc tctccccggag cccgcagcgg cccctgcggg gcccggcagc 180
 gcctctcgcg ggcacccgaaa ggcgcagccgc agggttctct accctcgagt ggtccggcgc 240
 cagctgccag tcgaggaacc gaacccagcc aaaaggcttc tctttctgtct gctaccatc 300
 gtcttctgcc agatcctgtat ggctgaagag ggtgtgccgg cgccctgtcc tccaaagagga 360
 cgccccctaac gcccgcattcc tgggcgccc cccctgtgtc cccctgcctc gagcccttta 420
 atctgacttc ggagccctcg gactacgctc tggacctcgact cactttcctc cagcaacacc 480
 cggccgcctt ctaactgtga ctccccgcac tccccaaaaaa gaatccgaaa aaccacaaag 540
 aaacaccagg cgtacccgtt ggcgcgagagc gtatccccaa ctgggacttc cgaggcaact 600
 tgaactcaga acactacagc ggagacgcca cccgggtgtt gaggcgggac cgaggcgac 660
 agagacccgag ggcgcatacgag accgaggcac agcccagctg ggggctaggc cgggtggaa 720
 ggagagcgtc gttttttttt ttcttattgc tccaaattaa tattttatgt tattttatgt 780
 cgtcctccca ggtgtatggag atgtgtacgt aatattttt ttaactttagt caagggtgtg 840
 agatgttccc cctgctgtaa atgcaggctt cttggatattt attgagcttt gtgggactgg 900
 tggaaagcagg acacccctggaa ctgcggcaaa gtaggagaag aaatggggag gactcgggtg 960
 ggggaggacg tcccggtgg gatgaagtct ggtgggtgggt cgtaagttta ggaggtgact 1020
 gcatccctca gcatctcaac tccgtctgtc tactgtgtga gacttcggcg gaccattagg 1080
 aatgagatcc gtgagatcct tccatcttct tgaagtcgccc tttagggtgg ctgcgaggta 1140
 gagggttggg ggttgggtgg ctgtcacgga ggcactgtcg agatcgctca gtatgttctg 1200
 tgaacacaaa taaaattgtat ttactgtctg ca 1232

<210> 16
 <211> 1678
 <212> DNA
 <213> Homo sapiens

<400> 16

gtcgccagga ggagcgcgcg ggcacagggt gcgcgtaccg aggcgtgcaa agactccaga 60
 attggaggca tggatgaagac tctgtgtgt tttgtggggc tgctgtgtac ctgggagagt 120
 gggcagggtcc tggggggacca gacgggtctca gacaatgagc tccaggaaat gtccaaatcag 180

ggaagtaagt acgtcaataa ggaaattcaa aatgtgtca acgggggtgaa acagataaaag 240
 actctcatag aaaaaacaaa cgaagagcgc aagacactgc tcagcaacct agaagaagcc 300
 aagaagaaga aagaggatgc cctaaatgag accaggaaat cagagacaaa gctgaaggag 360
 ctcccaggag tgtgaatga gaccatgatg gccctctggg aagagtgtaa gccctgcctg 420
 aaacagacct gcatgaagtt ctacgcacgc gtctgcagaa gtggctcagg cctgggtggc 480
 cgccagcttg aggagttcct gaaccagagc tcgccttct acttctggat gaatggtgac 540
 cgcatcgact ccctgctgga gaacgaccgg cagcagacgc acatgctgga tgtcatgcag 600
 gaccacttca gccgcgcgtc cagcatcata gacgagctct tccaggacag gttttcacc 660
 cgggagcccc aggataccta ccactacctg cccttcagcc tgccccacccg gaggcctcac 720
 ttcttccttc ccaagtcccg catcgccgc agctttgatg cccttctctc cgtacgagcc 780
 cctgaacttc cacgcccattt tccagccctt cttttagatg atacacgagg cttagcaggc 840
 catggacatc cacttccata gcccggcctt ccagcacccg ccaacagaat tcatacgaga 900
 aggacgat gacccggactg tgtgccccca gatccgcccac aactccacgg gctgcctgcg 960
 gatgaaggac cagtgtgaca agtgcgggaa gatcttgcgt gtgggactgt tccaccaaca 1020
 accccctccca ggctaaagctg cggcggggagc tcgacgaatc cctccaggc gctgagaggt 1080
 tgaccaggaa atacaacgag ctgctaaagt cctaccatgt gaagatgctc aacacccctt 1140
 ctttgcgtgga gcagctgaac gagcagtttta actgggtgtc ccggctggca aacccacgc 1200
 aaggcgaaga ccagtactat ctgcgggtca ccacgggtggc ttccccacact tctgactcgg 1260
 acgtttcccttc cgggtgtact gaggtggtcg tgaagcttt tgactctgtat cccatcactg 1320
 tgacggtccc tgtagaagtc tccaggaaga accctaaatt tatggagacc gtggcggaga 1380
 aagcgcgtca ggaataccgc aaaaagcacc gggaggagtg agatgtggat gttgttttg 1440
 cacctacggg ggcacatctgag tccagctccc cccaaagatga gctgcagccc cccagagaga 1500
 gctctgcacg tcaccaagta accaggcccc agccctccagg ccccaactc cgcccaaccc 1560
 ctccccgctc tggatcctgc actctaaccac tcgactctgc tgctcatggg aagaacagaa 1620
 ttgctcctgc atgcaactaa ttcaataaaa ctgtttgtg agctgaaaaa aaaaaaaaaa 1678

<210> 17
 <211> 1854
 <212> DNA
 <213> Homo sapiens

<400> 17

gtcttagttag ggacagacca agcacgcaaa acaaattgca atataatgtg ataagttctt 60
 taaaagaggt aagagcaacg tgctttgggaa gcagagaaga gggagaaagc agcatcttgc 120
 ctggatgagc caggggacac agaagagaag cccactatct catttaatct ttacaactct 180

cttgcagggt tccctgggtt gtgaaaatac atgagataaa tcatgaaggc cactatcatc 240
 ctccttctgc ttgcacaagt ttcctggggc tggaccgttt caacagagag gcttatttga 300
 ctttatgcta ggaagatgag gcttctgggg ataggcccag aagttcctga tgaccgcac 360
 ttcgagcccc tcccttagggc ccagtgtgcc cttccgctg tcaatgcacat cttcgagtgg 420
 tccagtgttc tgatttgggt ctggacaaag tgccaaagga tcttccccc gacacaactc 480
 tgcttagaccc gcaaaacaac aaaataaccg aaatcaaaga tggagacttt aagaacctga 540
 agaaccttca cgcattgatt cttgtcaaca ataaaattag gcaaagttag tcctggagc 600
 atttacaccc ttggtgaaag ttggAACGAC tttatctgtc caagaatcag ctgaaggaat 660
 tgccagaaaa aatgccccaa actcttcagg agctgcgtgc ccatgagaat gagatcacca 720
 aagtgcgaaa agttactttc aatggactga accagatgtat tgtcatagga actggcacc 780
 aatccgctga agagctcagg aattgaaaat ggggcttcc agggatgaa ggaagctctc 840
 ctacatccgc attgctgata ccaatatcac cagcattctt caaggtcttc ctcctccct 900
 tacgggaaatt acatcttgat ggcaacaaaa tcagcagagt tgatgcagct agcctgaaag 960
 gactgaataa ttggctaag ttgggattga gttcaacag catctctgtc gttgacaatg 1020
 gctctctggc caacacgcct catctgaggg agttcactt ggacaacaac aagcttacca 1080
 gagtacctgg tggctggca gaggcataagt acatccagggt tgtctacctt cataacaaca 1140
 atatctctgt agttggatca agtgacttct gcccacctgg acacaacacc aaaaaggctt 1200
 cttattcggg tgtgagtctt ttcagcaacc cggccagta ctgggagata cagccatcca 1260
 ctttcagatg tgtctacgtg cgctctgcca ttcaactcgg aaactataag taattctcaa 1320
 gaaagccctc attttataa cctggcaaaa tcttggtaat gtcattgcta aaaaataaaat 1380
 aaaagctaga tactggaaac ctaactgcaa tgtggatgtt ttacccacat gacttattat 1440
 gcataaaagcc aaatttccag tttaagtaat tgcctacaat aaaaagaaat tttgcctgcc 1500
 attttcagaa tcatctttt aagctttctg ttgatgttaa ctgagctact agagatattc 1560
 ttatccact aaatgtaaaa ttggagtaa atatatatgt caatatttag taaaagcttt 1620
 ctttttaat ttccaggaaa aaataaaaag agtatgagtc ttctgttaatt cattgagcag 1680
 tttagctcatt tgagataaag tcaaatgcca aacactagct ctgtttaat ccccatcatt 1740
 actggtaaag cctcatttga atgtgtgaat tcaatacagg ctatgtaaaa tttttactaa 1800
 tgtcattatt ttgaaaaat aaatttaaaa atacattcaa aattaaaaaa aaaa 1854

<210> 18
 <211> 1585
 <212> DNA
 <213> Homo sapiens
 <400> 18

gattcggcac gatggaatcc accagactaca tccagctccc tgaggcagag ttgagaatgg 60
 agagaatgtt acctctcctg actctgggc tcttggcggc tgggttctgc cctgctgtcc 120
 tctgccaccc taacagccca cttgacgagg agaatctgac ccagggagaa ccaagaccga 180
 gggacacacg tggacctcgg attagcctcc gccaacgtgg gacttcgctt tcagcctgta 240
 caagcagtta gtcctgaaag gcccctgata agaatgtcat cttctccca ctgaggcattc 300
 tccaccgcct tggccttcct gtctctggg gcccataat accaccctgg acagagattc 360
 tcaaaggcct caagttcaac ctcacggaga cttctgaggc agaaattcac cagagtttc 420
 cagcacctcc tgcgcaccct caatcagtcc agcgatgagc tgcaagctga gtatggaaa 480
 tgccatgttt gtcaaagagc aactcagtct gctggacagg ttcacggagg atgccaagag 540
 gctgtatggc tccgaggcct ttgccactga ctttcaggac tcagctgcag ctaagaagct 600
 catcaacgac tacgtgaaga atgaaactag gggaaaatc acagatctga tcaaggacct 660
 tgactcgcag acaatgatgg tccctggtaa ttacatcttc tttaaagcca aatggagat 720
 gccccttgac ccccaagata ctcatcagtc aaggttctac ttgagcaaga aaaaagtgggt 780
 aatggtgccc atgatgagtt tgcatcacct gactatacct tacttccggg acgaggagct 840
 gtcctgcacc gtggggcgc tgaagtacac aggcaatgcc agcgactct tcattccccc 900
 tgatcaagac aagatggagg aagtggaaagc catgctgctc ccagagaccc tgaagcggtg 960
 gagagactct ctggagttca gagagatagg tgagctctac ctgccaaagt tttccatctc 1020
 gagggactat aacctgaacg acataactct ccagctggc attgaggaag ctttcaccag 1080
 caaggctgac ctgtcaggga tcacagggc caggaaccta gcagtctccc aggtggtcca 1140
 taaggctgtg cttgatgtat ttgaggaggg cacagaagca tctgctgcc a cagcgtcaa 1200
 aatcaccctc ctttctgcat tagtggagac aaggaccatt gtgcgttca acaggccctt 1260
 cctgatgatc attgtccctt acagacaccc agaacatctt cttcatgagc aaagtccacca 1320
 atccccaaagca agcctagagc ttgccatcaa gcagtgggc tctcagtaag gaacttggaa 1380
 tgcaagctgg atgcctgggt ctctggcgc acgcctggccc ctgtgcaccc agtggccatg 1440
 gcatgtgtgg ccctgtctgc ttatccttgg aaggtgacag cgattccctg ttagctctc 1500
 acatgcacag gggcccatgg actcttcagt ctggagggtc ctgggcctcc tgacagcaat 1560
 aaataatttc gttggacacg ttaaa 1585

<210> 19
 <211> 1390
 <212> DNA
 <213> Homo sapiens

<400> 19

ggcaccacca ctaacctggg acagtgaatc gacaatgccg tcttctgtct cgtggggcat 60

cctccatgctg gcaggcctgt gctgcctggt ccctgtctcc ctggctgagg atccccaggg 120
 agatgctgcc cagaagacag atacatccca ccatgatcgat gatcacccaa ccttcaacaa 180
 gatcacccccc aacctggctg agttgcctt cagcctatac cgccagctgg cacaccagtc 240
 caacagcacc aatatcttct tctcccaagt gaggatcgct acagccttg caatgctctc 300
 cctgggggac caaggctgac actcacgatg aaatcctgga gggcctgaat ttcaacctca 360
 cggagattcc ggaggctcag atccatgaag gcttccagga actcctccgt accctcaacc 420
 agccagacag ccagctccag ctgaccacccg gcaatggcct gttcctcagc gagggcctga 480
 agcttagtggaa taagttttg gaggatgtta aaaagttgtta ccactcagaa gccttcactg 540
 tcaacttcgg ggacaccgaa gaggccaaga aacagatcaa cgattacgtg gagaagggt 600
 ctcaaggaa aattgtggat ttggtcaagg agcttgacag agacacagtt tttgctctgg 660
 tgaattacat cttctttaaa ggcaaatggg agagaccctt tgaagtcaag gacaccgagg 720
 aagaggactt ccacgtggac caggtgacca ccgtgaaggt gcctatgatg aagcgtttag 780
 gcatgtttaa catccagcac tgtaagaagc tgtccagctg ggtgctgctg atgaaatacc 840
 tggggcaatg ccaccgccc tttcttcctg cctgatgagg ggaaactaca gcacctggaa 900
 aatgaactca cccacgatat catcaccaag ttcctggaaa atgaagacag aaggtctgcc 960
 agcttacatt tacccaaact gtccattact ggaacctatg atctgaagag cgtcctgggt 1020
 caactggca tcactaaggt cttcagcaat ggggctgacc tctccgggt cacagaggag 1080
 gcaccctga agctctccaa ggccgtgcat aaggctgtgc tgaccatcga cgagaaaggg 1140
 actgaagctg ctggggccat gtttttagag gccataccca tgtctatccc ccccgaggc 1200
 aagttaaca aaccctttgt cttcttaatg attgaacaaa ataccaagtc tccctcttc 1260
 atggggaaag tggtaatcc cacccaaaaa taactgcctc tgcctcctca accctcccc 1320
 tccatccctg gccccctccc tggatgacat taaagaaggg ttgagctggt ccctgcctgc 1380
 atgtgactgt 1390

<210> 20
 <211> 1534
 <212> DNA
 <213> Homo sapiens

<400> 20

ggaagatccc aacagtttgc gccataaata taactttatc gcggacgtgg tggagaagat 60
 cgccccctgcc gtggttcata tcgaattgtt tcgcaagctt ccgtttctta aacgagaggt 120
 gccgggtggct agtgggtctg gttttattgt gtccgaagat ggactgatcg tgacaaatgc 180
 ccacgtggtg accaacaagc accgggtcaa agttgagctg aagaacggtg ccacttacga 240
 agccaaaaatc aaggatgtgg atgagaaagc agacatcgca ctcataaaaa ttgaccacca 300

<210> 21
<211> 2559
<212> DNA
<213> *Homo sapiens*

<223> unsure at all n locations
<400> 21

agctgtcgga gcggtagtt cgatttcgag cttggaggttt ccccccgcgc caggtgnact 60
tctcatcgct tgtttttctt ttgcatttt tttccacc gccgttgcgc ccctccccgt 120
cctggccgtc cgccctccgc cctctgcagg gacatctcta caccgttcccc atccggaaac 180
agggcaacat ctacaagccc aacaacaagg cttttggcaga cgagctgagc gagaagcaag 240
tgtacgacgc gcacaccaag gagatcgacc tggcaaccg cgaccctaaa cacctaaccg 300
atgacgtggc caagattgac tttgaagatg tgattgcaga accagaaggg acacacagtt 360

19

ttgacggcat ttgggaaggc cagcttcacc accttcaactg tgacgaaata ctggtttac 420
cgcttgctgt ctgcctctt tggcatcccg atggcaactca tctggggcat ttacttcgcc 480
attctctctt tcctgcacat ctggcagtt gtaccatgca ttaagagctt cctgattgag 540
attcagtgca tcagccgtgt ctattccatc tacgtccaca ccgtctgtga cccactctt 600
gaagctgttg ggaaaatatt cagcaatgtc cgcatcaact tgcagaaaga aatataaatg 660
acatttcaag gatagaagta tacctgattt ttttcctt taattttcctt ggtgccaatt 720
tcaagttcca agttgctaat acagcaacaa tttatgaatt gaattatctt ggtgaaaat 780
aaaaagatca ctttctcagt tttcataagt attatgtctc ttctgagcta tttcatctat 840
ttttggcagt ctgaattttt aaaacccatt taaattttt tccttacctt tttatggca 900
tgtggatcaa ccacatcgcttt attggctgag atatgaacat attgttggaa ggtaatttga 960
gagaaatatg aagaactgag gaggaaaaaa aaaaaaaaaaaga aaagaaccaa caacctaacc 1020
tgcctactcc aaaatgttgg tcattttatg ttaagggaaag aattccaggg tatggccatg 1080
gagtgtacaa gtatgtggc agatttcag caaactctt tcccactgtt taaggagttt 1140
gtggattact gccattcaact tcataatcca gtaggatcca gtgatcctta caagtttagaa 1200
aacataatct tctgccttct catgatccaa ctaatgcctt actcttcttgg aaatttttac 1260
ctatgatatt ttctgtgcct gaatattgt tatgttagata acaagacctc agtgccttcc 1320
tgttttcac attttcctt tcaaataggg tctaaactcag caactcgctt taggtcagca 1380
gcctccctga agacaaaaat tagaatatcc atgacctagt tttccatgcg tgtttctgac 1440
tctgagctac agagtctggt gaagctcaact tctgggcttc atctggcaac atctttatcc 1500
gtagtgggta tggttgacac tagccaaatg aaatgaatta aagtgggacc aatagggtcg 1560
agctctctgt gggctggca gtcctggaa gccagcttc cctgcctctc atcaactgaa 1620
tgaggtcage atgtctattt agcttcgttt atttcaaga ataatcacgc tttcctgaat 1680
ccaaactaat ccatcaccgg ggtgggttag tggctcaaca ttgtgttccc atttcagctg 1740
atcagtgggc ctccaaggag gggctgtaaa atggaggcca ttgtgtgagc ctatcagagt 1800
tgctgcaaacc ctgacccctg ctcaagtaaag cacttgcaac cgtctgttat gctgtgacac 1860
atggcccttc cccctgcccag gagcttggc cctaattccaa gcatctctt gcccagaaag 1920
aagatgggggg aggaggcagt aataaaaaaga ttgaagtatt ttgctggaat aagttcaaatt 1980
tcttctgaac tcaaactgag gaatttcacc tggtaaacctg agtcgtacag aaagctgcct 2040
ggtatatacca aaagctttt attcctcctg ctcatattgt gattctgcct ttggggactt 2100
ttcttaaacc ttcaagttatg atttttttt catacaactta ttggaaactct gcttgatttt 2160
tgccctttcc agtcttcctg acactttaat taccacccctg ttacctactt tgacttttg 2220
catttaaaac agacactggc atggatatacg ttttactttt aaactgtgtt cataactgaa 2280

20

aatgtgctat actgcatact ttttaaatgt aaagatattt ttatctttat atgaagaaaa 2340
 tcacttagga aatggctttg tgattcaatc tgtaaactgt gtattccaag acatgtctgt 2400
 tctacataga tgcttagtcc ctcatgcaaa tcaattactg gtccaaaaga ttgctgaaat 2460
 tttatatgct tactgatata ttttacaatt ttttatcatg catgtcctgt aaaggttaca 2520
 agcctgcaca ataaaaatgt ttaacggtta aaaaaaaaaa 2559

<210> 22
 <211> 981
 <212> DNA
 <213> Homo sapiens

<400> 22

gcggagtc tc caactgggag agctgcagct gccgagagga ggagaacgct gaggtcggtc 60
 ggaccaacgg acgcgcgtgac cgctgccaac tgca gctgccttc 120
 tgccactaag gtatgtccgccc tttctatgag ccctccccaa gattagctgg gtgcggggtg 180
 gtggagccg ttctttggtg gctgaagccc ctctcctgct gctcctcctg caggtcactc 240
 ccgcctccga gagccccagag ccgagatgga aacggtccag gagctgatcc ccctggccaa 300
 ggagatgatg gcccagaagc gcaaggggaa gatggtaag ctgtacgtgc tggggcagcg 360
 tgctggccct ctgcggcgtg gtgcgtccgccc tgatggagac tgggtgcagc cccttcacgg 420
 ccgcacagacg tctgcgggac caggaggcag ccgtggcggaa gctgcaggcc gcccgggac 480
 gacaggctct ccagaagcaa gccctgcagg agaaaggcaa gcagcaggac acggcctcg 540
 gcggccgggc cctgtccaaac cggcagcacg cctccttagga actgtgggag accagcggag 600
 tgggagggag acgcagttaga cagagacaga ccgagaagga agggagagac agagggggcg 660
 cgcgcacagg agcctgactc cgctgggaga gtgcaggagc acgtgctgtt ttttattttg 720
 acttaacttc agagaaaccg ctgacatcta gaactgaccc accacaagca tccaccaaag 780
 gagtttggga ttgatgttttgc tgcgttgca gcactgcatt gtcatgacat ttccaaact 840
 gtgtgaatta tctaaatgct tctaccattt tgcaactaggg aggaaggata aatgttttt 900
 atgttattat tattaattat tacaatgacc accatgtgc attttggaaat aaaaaacttt 960
 ttataccaaa aaaaaaaaaa a 981

<210> 23
 <211> 835
 <212> DNA
 <213> Homo sapiens

<400> 23

gcactccaa agaactgggt actcaacact gaggcagatc ttttttttga ggctaaaaac 60
 catgtgctgtt accaagagtt tgcgtccggg ctgttttgc gtcagtgcgtg ctactccacc 120

21

tctgcggcga atcagaagca gcaagcaact ttgactgctg tcttggata cacagaccgt 180
 attcttcattc ctaaatttat tgtggcttc acacggcage tggccaatga aggctgtgac 240
 atcaatgcta tcatactttca cacaagaaaa aagttgtctg tgtgcgaaa tccaaaacag 300
 acttgggtga aatatattgt gcgtctcctc agtaaaaaag tcaagaacat gtaaaaactg 360
 tggctttctt ggaatggaat tggacatagc ccaagaacag aaagaacctt gctggggttg 420
 gaggtttcac ttgcacatca tggagggttt agtgcttatac taaaaattgtgc ctcactggac 480
 ttgtccaatt aatgaagttt attcatattt catcatagtt tgctttttt aagcatcaca 540
 ttaaagttaa actgtatattt atgttattta tagctgtagg ttttctgtgt ttagctattt 600
 aatactaatt ttccataagc tattttggtt tagtcaaag tataaaatta tatttggggg 660
 ggaataagat tataatggact ttcttgcaag caacaagcta tttttaaaaa aaaactattt 720
 aacattcttt tgtttatattt gtttgcctc ctaaattgtt gtaattgcat tataaaataa 780
 gaaaaatattt aataagacaa atattgaaaa taaagaaaaca aaaagttcaa aaaaa 835

<210> 24
 <211> 981
 <212> DNA
 <213> Homo sapiens

<400> 24

gcgcggcggaa gagctcttgc gcgtcttgc tttgcctgggt gtcgggtggtt agtttctgctg 60
 acttgggttg ggactgctga taggaagatg tcttcaggaa atgctaaaaat tggcaccct 120
 gcccccaact tcaaagccac agctgttatg ccagatggtc agtttaaaga ttcagccctg 180
 tctgactaca aaggaaaata tgggtgttcc ttcttttacc ctcttgactt cacctttgtg 240
 tgccccacgg agatcattgc ttctcatttgc ttcattttacc ctcttgactt cacctttgtg 300
 ccaagtgatt ggtgcttctg tgggattctc acttctgtca tctagcatgg ggtcaataaca 360
 cctaagaaac aaggaggact gggaccatg aacattcattt tggatcaga cccgaagcgc 420
 accattgctc aggattatgg ggtcttaaag gctgtatgaa gcatctcggtt cagggccctt 480
 ttatcatttgc atgataaggg tatttttcgg cagatcactg taaatgaccc ccctgttggc 540
 cgctctgtgg atgagacttt gagactagtt caggcccttcc agttcactga caaacatggg 600
 gaagtgtgcc cagctggctg gaaacctggc agtgatacca tcaagcctga tgtccaaaag 660
 agcaaagaat atttctccaa gcagaagtga gcgctggct gtttttagtgc caggctgcgg 720
 tgggcagcca tgagaacaaa acctcttctg tattttttt ttccatttagt aaaacacaag 780
 acttcagatt cagccgaatt gtgggtgtttt acaaggcagg ctttcctac aggggggtgga 840
 gagaccagcc ttcttcctt tggtaggaat ggctgagtt ggcgttgtgg gcaggctact 900
 gttttgtatg atgtatttagt agagcaaccc attaatcttt ttagttgtt attaaacttg 960

aactgagaaa aaaaaaaaaa a

981

<210> 25
 <211> 1642
 <212> DNA
 <213> Homo sapiens

<400> 25

aaaaaaaaaggcg agcccgcccc ccctggagac cccggctctca cggagttgac gtcatgacct 60
 acgtgaggga gacctgcggg tgctgcact gtgagaagcg ctgtggcgcc ctggacgtgg 120
 tcttcgtcat cgacagctcc gagagcattg ggtacaccaa cttcacactg gagaagaact 180
 tcgtcatcaa cgtggtcaac aggctgggtg ccatcgctaa ggaccccaag tccgagacag 240
 ggacgcgtgt gggcgtggtg cagtacagcc acgagggcac ctttgaggcc atccagctgg 300
 acgacgaaca tatcgactcc ctgtcgagct tcaaggaggc tgtcaagaac ctcgagtgg 360
 ttgcggccgg cacctggaca ccctcagccc tcaagtttgc ctacgaccgc ctcatcaagg 420
 agagccggcg ccagaagaca cgtgtgttg cgggtggtcat cacggacggg cgccacgacc 480
 ctcgggacga tgacctaacc ttgcggccgc tgtgcgaccg cgacgtcaca gtgacggcca 540
 tcggcatcggtt ggacatgttc cacgagaagc acgagagtga aaacctctac tccatgcct 600
 ggcacaagcc acagcagggtg cgcaacatga cgctgttctc ccgacctggg cgggttagaa 660
 gttcatcgat gacatgggag gacgtcctct gcccggaccc tcagatcgtg tgcccagacc 720
 ttccctgcca aacagagctg tccgtggcac agtgcacgca gcccggcgtg gacatcgat 780
 tcctgctgga cggctccgag cggctgggtg agcagaactt ccacaaggcc cggcgcttcg 840
 tggagcaggt ggcgcggccgg ctgacgctgg cccggaggga cgacgacccct tcacacgcac 900
 gcgtggcgct gctgcagtt ggtggccccc gcgagcagca ggtggccttc ccgctgagcc 960
 acaacctcac ggcacatccac gaggcgctgg agaccacaca atacctaacc tccttcgc 1020
 acgtggccgc aggcggtggt cacgccatca atgcctatgt ggcgcggccag cgtggccggcc 1080
 ggcggaggca cgcagagctg tccttcgtgt tcctcacgga cggcgctcag ggcaacgaca 1140
 gtctgcacga gtcggccgcac tccatgcgc aacgaaacgt ggtacccacc gtgctggcc 1200
 tggggcagcga cgtggacatg gacgtgctca ccacgctcag cctgggtgac cgtggccgg 1260
 tggccacga gaaggactat gacagcctgg cgcaacccgg ctcttcgac cgcttcatcc 1320
 gctggatctg ctacgcggccgc cggccggcc cccggcgtcga gggtcgtgag cccacccct 1380
 ccatggtgct aagcgcccccc gggcccccaca cggccagcac cgctgcctac tcggacgacg 1440
 ccctggccct gcacccctcc agctccccc acggggtccc cgtagccccc gccccggccc 1500
 agccccaggt ctccccaggg cctccgcagg ctggccggcc tcctccccc tgcagccatc 1560
 ccaaggctcc tgacctaccc gggccctgag ctctggagca agccctgacc caataaaggc 1620

tttgaaccca aaaaaaaaaa aa

<210> 26
 <211> 163
 <212> DNA
 <213> Homo sapiens

<400> 26

gaccagttt tcaagaagg tagctgtgg agggggacac accctctgtc tgatccctta 60
 tcaaagagga caaggaaact atagagctga ttttagaata ttttacaaat acatgccttc 120
 cattggaatg ctaagatttt ctactgttc tggggacggg aaa 163

<210> 27
 <211> 1746
 <212> DNA
 <213> Homo sapiens

<223> unsure at all n locations
 <400> 27

cagcgctccc actctcgccc gacacccttc atggccaacc gttacaccat ggatctgact 60
 gccatctacg agagcctcct gtcgctgagc cctgacgtgc ccgtgccatc cgaccatgga 120
 gggactgagt ccagcccagg ctggggctcc tcgggaccct ggagcctgag cccctccgac 180
 tccagccctgt ctggggtcac ctccccctg cctggccgct ccaccagcct agtggagggc 240
 cgcagctgtg gctgggtgcc cccacccct ggcttcgcac cgctggctcc ccgcctggc 300
 cctgagctgt caccctcacc cacttcgccc actgcaacct ccaccacccc ctgcgcgtac 360
 aagactgagc tatgtcggac cttctcagag agtgggcgtc gcccgtacgg ggccaagtgc 420
 cagtttgcctt atggcctggg cgagctgcgc cagggcaatc gccacccaa atacaagacg 480
 gaactctgtc acaagttcta cctccagggc cgctgcccct acggctctcg ctgccacttc 540
 atccacaacc ctagcgaaga cctggggccc ccggggccacc ctccctgtgt tcgcccagagc 600
 atcagcttct ccggcctgcc ctctggccgc cggacctcac caccaccacc aggcctggcc 660
 ggcccttccc tgcctccag ctccctctcg ccctccagct cccaccacc acctggggac 720
 ctccactgt naccctctgc cttctctgtc gcccctggca ccccccctggc tcgaagagac 780
 cccaccccaag tctgttgcctt ctccctgccga agggccactc ctatcagcgt ctggggggccc 840
 ttgggtggcc tggttcggac cccctctgtc cagttccctgg ggatccgacc ctgatgaata 900
 tgccagcagc ggcagcagcc tggggggctc tgactctccc gtcttcgagg cgggagttt 960
 tgcaccaccc cagccccgtgg cagccccccg gcgactcccc atcttcaatc gcatctctgt 1020
 ttctgagtga caaagtgact gcccggtcag atcagctggc tctcagcggg gagccacgtc 1080
 tcttgcaactg tggtctctgc atggacccca gggctgtggg gacttggggg acagtaatca 1140
 agtaatcccc ttttccagaa tgcatttaacc cactccccctg acctcacgtc ggggcaggc 1200

cccaaagtgtg caagctcagt attcatgatg gtggggatg gagtgtcttc cgaggttctt 1260
 gggggaaaaaa aaattttagc atatttaagg gaggcaatga accctctccc ccacctcttc 1320
 cctgccccaaa tctgtctctt agaatctt atgtgtgtgaa taataggcct tcactgcccc 1380
 tccagttttt atagacctga ggttccagtg tccctggta actggAACCT ctccctgaggg 1440
 ggaatcctgg tgctcaaatt accctccaaa agcaagtagc caaagccgtt gccaaacccc 1500
 acccataaaat caatgggccc ttattttatg acgactttat ttattctaat atgattttat 1560
 agtatttata tatattgggt cgtctgcttc ccttgttattt ttcttccttt ttttgtata 1620
 ttgaaaacga cgatataatt attataagta gactataata tatttagtaa tatataattat 1680
 taccttaaaa gtctatttt gtgtttggg catttttaaa taaacaatct gagtgtaaaa 1740
 aaaaaaa 1746

<210> 28
 <211> 1884
 <212> DNA
 <213> Homo sapiens

<400> 28

cgtcgtagcc ccaacctcgaa cggtcgcccgt gggccgggtc gcgtctgcct tggagaagaa 60
 gacaaagagc aagggggccct acatctgcgc tctgtgcgc aaggagttca agaacggcta 120
 caatctccgg aggcacgaag ccattccacac gggagccaag gccggccggg tccctcggg 180
 tgctatgaag atgcccacca tggtgccccc gaggctccgt agcgtgcctt agctgagcgg 240
 agccggcggg ggagggggag aggcgggtgc cggcggcggc gctgcccggc tggccgggg 300
 tggcgtggtg accacgaccc cctcgaaaa ggcacatccgg aagaaccatg cctgcgagat 360
 gtgtggcaag gccttcggcg acgtctacca cctgaaccga cacaagctgt cgcaactcgga 420
 cgagaagccc taccagtgcc cggtgtgcca gacgcgttc aagcgcaagg accgcattgag 480
 ctaccacgtg cgctcacatg acggcgctgt gacaaagccc tacaactgtt cccactgtgg 540
 caagagcttc tcccgccgg atcacctcaaa cttcacgtc agacaagtgc actcaacaga 600
 acggcccttc aaatgtgaga aatgtgaggc aatgtgaggc acgaaggatc ggctgcccc 660
 gcacacagta cgacacgagg agaaagtgc aatgtgaggc tggggcaaga tgctgagctc 720
 ggcttatatt tcggaccaca tgaaggtgca cagccagggt cctcaccatg tctgtgagct 780
 ctgcaacaaa ggtactggtg aggtttgtcc aatggcggcg gacgcggcag cggccggca 840
 gcccccccttc cggcagcgt agcagccctt cccacagctg tgggtccctt ctcggggggcg 900
 gaggggggtgc ctgtgagctc tcaagccactt ccccccac cctggtgagc tccaagttgg 960
 ttgcggggga gagggggagaa tggagtagag tcccttggta caagctccctc tccccctct 1020
 tttccacca actcctatcc ccttaccaac caaggagctt ccagaaggaa aggaggaaga 1080

aatgttttct taggggaatt cgcttaggttt taacgatttg tttctcctgc tcctcttcta 1140
 tcagacctga ccccacacaa acctgtcccc tgggtgtgt tgaagtcccc tggacagtgg 1200
 gcaggggtgg cagaggacac gagcagccac tgcccgtaacc ccctctcctc tctgttaagcc 1260
 catgccctgt ctcccaggg acttgtgagc ctctccctc gacggtcctc ttctctcctt 1320
 ccagtcctct cccccctgctg tctgcagccc ctccccgggg agttggtgct ttctttctt 1380
 tttttttttt ttccagggg gagggaggag aggaaggagg gggatcagag ctgtcccaaa 1440
 gagggaaaagc ggtgaggtt gaggagggc agaagcaggg ccggcaaaagg ttgtaccttc 1500
 ataagggtggt atgggggggtt ggggtcaggc cctgaacatc gtcctacttg agaatctgtc 1560
 aggggaaaaaa gtcaggggaa gcaggaggaa gagccaggag gccagaggca gagaagagat 1620
 ggagtcttag gggccagggt gagcgagggg tccagggctt agaggtgttt cctggggggcg 1680
 ggggaatgca gccagtgtcc ccctccctc ttccacccca gctccagccc tggctttgtc 1740
 ttttcatccc tctcccac gacagaagaa gttgtggccc tggccatgtc atcggttcc 1800
 tgggtccctt gcatgtaccc caccctccac cccttccttt tgccggacc ccattacaat 1860
 aaattttaaa taaaatcctg aaaa 1884

<210> 29
 <211> 1563
 <212> DNA
 <213> Homo sapiens
 <400> 29

tcacctccag. gatacagaca gccccccctc agcccagccc agccagggtct cctacacccgc 60
 caccatgcca ttccggtaaca cccacaacaa gttcaagctg aattacaagc ctgaggagga 120
 gtaccccgac ctcagcaaac ataacaacca catggccaag gtactgaccc ttgaactcta 180
 caagaagctg cgggacaagg agactccatc tggcttact gttagacgatg tcacccagac 240
 aggagtggac aaccctggc accccttcat catgaccgtg ggctgcgtgg ctggtgatga 300
 ggagtctac gaagtttca aggaactctt tgacccatc atctcgatc gacccgggg 360
 ctacaaaccc acttgacaag cacaagactg acctcaacca ttgaaaacct caagggtgga 420
 gacgacccctt acccttaacta cgtgctcagc agccgcgtcc gcactggcc cagcatcaag 480
 ggctacacgt tgccccccaca ctgctccctg ggcgagcgcc gggcggtggaa gaagctctt 540
 gtggaaagctc tcaacagccct gacggggcgag ttcaaaaggga agtactaccc tctgaagagc 600
 atgacggaga aggaggcagca gcagctcatc gatgaccact tccctgttgcgca caagcccggt 660
 tccccctgc tgctggcctc aggcatggcc cgccactggc ccgacgcccc tggatctggc 720
 acaatgacaa caagagcttc ctgggtgtgg tgaacgagga ggatcaccc cgggtcatct 780
 ccatggagaa gggggggcaac atgaaggagg tttccgcgg ctctgcgtt gggctgcaga 840

agattgagga gatcttaag aaagctggcc accccttcat gtggaaccag cacctggct 900
 acgtgctcac ctgcccattcc aacctggca cctgggctgc gtggaggcgt gcatgtgaag 960
 cctggcgac ctgagcaagc accccaagtt cgaggagatc ctcaccgc tgcgtctgca 1020
 gaagaggggt acaggtggcg tggacacagc ctgccgtggg ctcagtattt gacgtgtcca 1080
 acgctgatcg gctgggctcg tccgaagtag aacaggtgca gctggtggtg gatggtgtga 1140
 agctcatggt gaaaaatggag aagaagttgg agaaaggcca gtccattgac gacatgatcc 1200
 ccccccagaa gtaggcccct gcccacctgc caccgactgc tggaaaccag ccagtggag 1260
 ggcctggccc accagagtcc tgctccctca ctctctggccc cggccctgt cccagagtcc 1320
 cacctggggg ctctctccac ctttcaga gttccagttt caaccagagt tccaaccaat 1380
 gggctccatc ctctggattc tggccaatga aatatctccc tggcagggtc ctcttcttt 1440
 cccagagctc caccccaacc aggagctcta gttaatggag agctcccagc acactcggag 1500
 cttgtgttt gtctccacgc aaagcgataa ataaaagcat tggtggcctt aaaaaaaaaa 1560
 aaa 1563

<210> 30
 <211> 2263
 <212> DNA
 <213> Homo sapiens

 <223> unsure at all n locations
 <400> 30

ctcgagacaa gcccgtatgt gtcaacacct atggaagcta caggtgccgg accaacaaga 60
 agtgcagtcg gggctacgag cccaaacgagg atggcacage ctgcgtgggg actctcgccc 120
 agtcacccggg cccccggccc accnnnnnnna cncccccggac cggggctggg agcaagcagg 180
 cggcggcgcc ggcggcagag gcccgcgcgaa gcccgcgtt cccacgcccc tagggggcgg 240
 gggccgagagc gggaggatgg ctccgagcgc tgacccggc atgtccagga ttttaccgtt 300
 cctgctgctg ctctggtttc tgcccatcac tgaggggtcc cagcggctg aacccatgtt 360
 cactgcagtc accaactcag ttctgcctcc tgactatgac agtaatccca cccagctcaa 420
 ctatggtgtg gcagttactg atgtggacca tgatggggac tttgagatcg tcgtggcggg 480
 gtacaatgga cccaaacctgg ttctgaagta tgaccgggccc cagaaggccc tggtgaacat 540
 cgcggtcgt gaggcgtacta acccctacta cgcgtgcgg gaccggcagg ggaacgccc 600
 cggggtcaca gcctgcgaca tcgacgggaa cggccgggag gagatctact tcctcaacac 660
 caataatgcc ttctcgaaaa tggccacgta caccgacaag ttgttcaagt tccgcaataa 720
 ccgggtggaa gacatcctga gcgatgaggt caacgtggcc cgtggtgtgg ccagccttt 780
 tgccggacgc tctgtggcct gtgtggacag aaagggtct ggacgtact ctatctacat 840

27

tgccaaattac gcctacggta atgtgggccc tcatgcctc attgaaatgg accctgaggc 900
 cagtgacctc tcccgccccca ttctggcgct cagagatgtg gctgctgagg ctggggtcag 960
 caaatataca gggggccgag gcgtcagcgt gggccccatc ctcagcagca gtgcctcgga 1020
 tatcttctgc gacaatgaga atgggcctaa ctccctttc cacaaccggg gcgatggcac 1080
 ctttggac gctgcggcca gtgctgggtgt ggacgacccc caccagcatg ggcgaggtgt 1140
 cgccctggct gacttcaacc gtgatggcaa agtggacatc gtctatggca actggaatgg 1200
 cccccaccgc ctctatctgc aaatgagcac ccatgggaag gtccgcttcc gggacatcgc 1260
 ctcacccaag ttctccatgc cctccctgt ccgcacggtc atcaccgccc actttgacaa 1320
 tgaccaggag ctggagatct tcttcaacaa cattgcctac cgcaagctc cagccaaccc 1380
 cctctccgc gtcatccgta gagagcacgg agacccctc atcgaggagc tcaatccgg 1440
 cgacgccttg gagcctgagg gccggggcac aggggggtgtg gtgaccgact tcgacggaga 1500
 cgggatgctg gacctcatct tgtcccatgg agagtccatg gctcagccgc tgtccgtctt 1560
 cccgggcaat cagggcttca acaacaactg gctgcgagtg gtgccaacgc acccggttg 1620
 gggccttgc caggggagct aaggtcgtgc tctacaccaa gaagagtggg gcccacctga 1680
 ggatcatcga cgggggctca ggctacctgt gtgagatgga gcccgtggca cactttggcc 1740
 tggggaaagga tgaagccagc agtgtggagg tgacgtggcc agatggcaag atggtaggcc 1800
 ggaacgtggc cagcggggag atgaactcag tgctggagat cctctacccc cgggatgagg 1860
 acacacttca ggacccagcc ccactggagt gtggccaagg attctccag cagaaaaatg 1920
 gccattgcca tggacaccaa tgaatgcattc cagttcccat tcgtgtggcc tcgagacaag 1980
 cccgtatgtg tcaacaccta tggaaacctac aggtgcccga ccaacaagaa gtgcagtcgg 2040
 ggctacgagc ccaacgagga tggcacagcc tgcgtggctc aagtggccctt tttaggtggg 2100
 tattcttcag cccgcctctag aatctctgag cctctctctc gggcctcata tctttctcta 2160
 ggccttggac tttgccttca gttatatgca ctttaaatcc catcaataaa ggaaaaaaaca 2220
 aaacaaaaact aacagcctt gtggaaaact aaaaaaaaaaaa aaa 2263

<210> 31
 <211> 2310
 <212> DNA
 <213> Homo sapiens

<400> 31

cggcattcct cctgttagctg cacgaagcac cttggaaagtt gtttcaacc atatccagcc 60
 tttgccaaat acatcctatc tgccacacat ccagcgtgag gtccctccag ctacaaggtg 120
 ggcaccatgg cggagaagtt tgactgccac tactgcaggg atcccttgca ggggaagaag 180
 tatgtgcaaa aggtggcca ccactgctgc ctgaaatgct ttgacaagtt ctgtgccaac 240

28

acctgtgtgg aatgccgcaa gcccacatcggt gcggactccca aggaggtgca ctataagaac 300
cgcttctggc atgacacacctg cttccgctgt gccaagtgcc ttcacccctt gggccaatga 360
gacctttgtg gccaaggaca acaagatcct gtgcaacaag tgcaccactc gggaggactc 420
ccccaaagtgc aaggggtgct tcaaggccat tgtggcagga gatcaaaaacg tggagtacaa 480
ggggaccgtc tggcacaaaag actgcttcac ctgttagtaac tgcaagcaag tcatacgac 540
tggaagcttc ttccctaaag gggaggactt ctactgcgtg acttgccatg agaccaagtt 600
tgccaagcat tgcgtgaagt gcaacaaggc catcacatct ggaggaatca cttaccagga 660
tcagccctgg catgcccattt gctttgtgtg tgttacctgc tctaagaagc tggctggca 720
gcgttccacc gctgtggagg accagtattt ctgcgtggat tgctacaaga actttgtggc 780
caagaagtgt gctggatgca agaaccat cactgggttt ggtaaaggct ccagtgtggt 840
ggcctatgaa ggacaatcct ggcacgacta ctgcttccac tgcaaaaaat gctccgtgaa 900
tctggccaac aagcgctttg tttccacca ggagcaagtg tattgtcccg actgtgcca 960
aaagctgtaa actgacaggg gtcctgtcc tggtaaatgg catttgaatc tgcgtctttg 1020
tgccttact ttctgcctta taccatcaat agggaaagag tggtccttcc cttctttaaa 1080
gttctcccttc cgtctttct cccattttac agtattactc aaataaggc acacagtgtat 1140
catattagca ttttagcaaaa agcaaccctg cagcaaaagtg aatttctgtc cggctgcaat 1200
ttaaaaatga aaacttaggt agattgactc ttctgcatttgc ttctcataga gcagaaaaatg 1260
gctaattcatt tagccactta gtgatgtaa caagaagcat aggagataaa acccccactg 1320
agatgcctct catgcctcag ctgggaccca cccgtgtaga cacacgacat gcaagagttg 1380
cagcggctgc tccaaactcac tgctcacccct cttctgtgag caggaaaaga accctactga 1440
catgcattgtt ttaacttcct catcagaact ctgccttcc ttctgttctt ttgtgtttc 1500
aaataactaa cacgaacttc cagaaaatta acatttgaac ttagctgtaa ttctaaactg 1560
acctttcccc gtactaacgt ttggtttccc cgtgtggcat gtttctgtgag cgttcctact 1620
ttaaagcatg gaacatgcag gtgatttggg aagtgttagaa agacctgaga aaacgagcct 1680
gtttcagagg aacatcgtca caacgaatac ttctggaaagc ttaacaaaac taaccctgct 1740
gtccttttta ttgtttttaa ttaatatttt tggatattt gatagcaaaa tagtttatgg 1800
gtttggaaac ttgcattgaaa atattttagc cccctcagat gtcctgcag tgctgaaatt 1860
catcctacag aagtaaccgc aaaactctag agggggagtt gaggcaggcgc cagggctgac 1920
atcaacatgg atatgacatt tcacaacagt gactagttga atcccttgcatacgttagt 1980
tgtctgcctt ttgtccatgt gttatgagg actgcaaatg cccttctgtt gtgattccta 2040
ggactttcc tcaagaggaa atctggattt ccacccatcg cttacctgaa atgcaggatc 2100
acccatctac tgcatttttttattt gacatgtat gatggagacaa tatcaaaaatg 2160

29

aaacatgtaa tgacaataca tactaacatt cttgttaggag tggtagaga agctgatgcc 2220
tcatttctac attctgtcat tagctattat catctaacgt ttcagtgtat ccttacagaa 2280
ataaagcagc atatgaataa aaaaaaaaaa 2310

<210> 32
<211> 3342
<212> DNA
<213> Homo sapiens

<400> 32

gaagaagtta agagcttcat ggatcgaaag aagggattta cagaagttaa gtcgcagaat 60
ggagaattca tgacccacaa acttaaacat actgagaata ctttcagccg ccctggaggg 120
agggccagcg tggacaccaa ggaggcttag ggcgcggggc aggtggaaagc cggcaaaagg 180
ctggaggagc ttcgtcgtag tcgcggggag accgagagcg aagagttcga gaagctcaa 240
cagaagcagc aggaggcggc tttggagctg gaggactca agaaaaagag ggaggagaga 300
aggaagggtcc tggaggagga agagcagagg aggaagcagg aggaagccga tcgaaaactc 360
agagaggagg aagagaagag gaggctaaag gaagagattg aaaggcgaag agcagaagct 420
gctgagaaac gccagaagat gccagaagat ggcttgcag atgacaagaa accattcaag 480
tgttcaactc ctaaagggttc atctctcaag atagaagagc gagcagaatt tttgaataag 540
tctgtgcaga aaagcagtgg tgtcaaatcg acccatcaag cagcaatagt ctccaagatt 600
gacagcagac tggagcagta taccagtgcatttggagggaa caaaaagcgc aaaaacctaca 660
aagccggcag cctcgatct tcctgttct gctgaaggtg tacgcaacat caagagtatg 720
tgggagaaag ggaatgtgtt ttcatcccc actgcagcag gcacacaaaa taaggaaact 780
gcctggctt aaggtagggg ttcttagccg catcaatgaa tggctaacta aaacccaga 840
tggaaacaag tcacctgctc ccaaacccttc tgacttgaga ccaggagacg tatccagcaa 900
gcggAACCTC tggaaaagc aatctgtgga taaggtcact ttccccact aaggttttag 960
acagttccag aaagaaccca agctcaagac gcaggacgag ctcagttgta gagggctaat 1020
tcgctctgtt ttgtattttt gttgatttac taaattgggt tcattatctt ttatTTTCA 1080
atatcccagt aaacccatgt atattatcac tatatttaat aatcacagtc tagagatgtt 1140
catggtaaaa gtactgcctt tgcacaggag cctgtttcta aagaaaccca tgctgtgaaa 1200
tagagacttt tctactgatc atcataactc tgtatcttag cagtgataacc aaccacatct 1260
gaagtcaaca gaagatccaa gtttaaaatt gcctgcggaa tgtgtgcagt atctagaaaa 1320
atgaaccgta gttttgttt ttttaaatac agaagtcatg ttgttctgc actttataat 1380
aaagcatgga agaaatttac ttagtaggca attgtAACAC tttttgaaag taacccattt 1440
cagatttggaa atactgcaat aatggttgtc tttaaaaaaaaaaaaaaaaat gtactgttaa 1500

30

3342

<211> 954
 <212> DNA
 <213> Homo sapiens

<400> 33

cagcctcaag attcacagca tctcagacgc agcctaggcc gcaccaggat gtcggacacc 60
 gaggaggcagg aatatgagga ggagcagccg gaagaggagg ctgcgggtga ggaggaggaa 120
 gcccccaag agccggagcc ggtggcagag ccagaagagg aacgccccaa accaagccgc 180
 cccgtggtgc ctcccttgc cccgccaaag atcccagaag gggagcgcgt tgacttcgt 240
 gacatccacc ggcaagcgcga tggagaaaga cctgctggag ctgcagacac tcacatcgatgt 300
 acatttcgag cagcggaaaga aggaggaaga ggagctggtt gccttgaagg agcgcattga 360
 gcgccgccc tcagagagag cccgagcaac agcgcttcag aactgagaag gaacgcgaac 420
 gtcaggctaa gctggcggag gagaagatga ggaaggaaga ggaagaggcc aagaagcggg 480
 cagaggatga tgccaaagaaa aagaaggtgc tgcacat gggggcccat tttggcggct 540
 acctggtcaa ggcagaacag aagcgtggta agcggcagac ggggcgggag atgaaggtgc 600
 gcatccctc cgagcgtaaag aagcctctgg acattgacta catgggggag gaacagctcc 660
 gggagaaagc ccaggagctg tcggactgga tccaccagct ggagtctgag aagttcgacc 720
 tgcgtggcgcgaa gctgaaacag cagaaatatg agatcgacgt gctgtacaac cgcacatcgcc 780
 acgcccagaa gttccggaaag gggcagggaa agggccgcgt tggaggccgc tggaaagttag 840
 gatgccgccc cggacagtgg cacctggaa gcctgggagt gtttgcctca tggtagctt 900
 gaaaataaacg ctccccctcag acacccgcgtg gtttctctga tggattatgt gtttgcgtt 954

<210> 34
 <211> 3183
 <212> DNA
 <213> Homo sapiens

<400> 34

gcgcccgcacc tacaccagcc aacccagatc ccgaggtccg acagcgcggc gcccagatcc 60
 ccacgcctgc caggagcaag ccgagagcca gccggccggc gcactccgac tccgagcgt 120
 ctctgtcctt cgacccgagc cccgcgcctt tccgggacc cctgccccgc gggcagcgcgt 180
 gccaacctgc cggccatgga gacccgtcc cagcggcgcg ccacccgcag cggggcgcag 240
 gccagctcca ctccgctgtc gcccacccgc atcaccggc tgcaggagaa ggaggacctg 300
 caggagctca atgatcgctt ggcggctac atcgaccgtg tgcgtcgct ggaaacggag 360
 aacgcagggc tgcgccttcg catcaccgag tctgaaggagg tggtcagccg cgaggtgtcc 420
 ggcacatcaagg ccgcctacga ggccgagctc gggatgccc gcaagacccct tgactcagta 480
 gccaaggagc gcgccgcct gcagctggag ctgagcaaag tgcgtgagga gttttaaggag 540

32

ctgaaagcgc ggcaatacca agaaggaggg tgacctgata gctgctcagg ctcggctgaa 600
ggacctggag gctctgctga actccaagga ggccgcactg agcactgctc tcagtgagaa 660
gcccacgctg gagggcgcagc tgcatgatct gcggggccag gtggccaagc ttgaggcagc 720
cctaggtgag gccaagaagc aacttcagga tgagatgctg cggcgggtgg atgctgagaa 780
caggctgcag accatgaagg aggaactgga cttccagaag aacatctaca gtgaggagct 840
gcgtgagacc aagcgccgtc atgagacccg actgggtggag attgacaatg ggaagcagcg 900
tgagtttgag agccggctgg cggatgcgtc gcaggaactg cgggcccagc atgaggacca 960
ggtggagcag tataagaagg agctggagaa gacttattct gccaagctgg acaatgccag 1020
gcagtctgct gagaggaaca gcaacctggc gggggctgcc cacgaggagc tgcaagcagtc 1080
gcgcacatccgc atcgacacgac tctctgccc gctcagccag ctccagaagc agctggcagc 1140
caaggaggcg aagtttcgag acctggagga ctcactggcc cgtgagcggg acaccagccg 1200
gcggctgcct ggccggaaaag gagcggggaga tggccgagat gcgggcaagg atgcagcagc 1260
agctggacga gtaccaggag cttctggaca tcaagctggc cctggacatg gagatccacg 1320
cctaccgcaa gctcttggag ggcgaggagg agaggctacg cctgtccccc agccctacct 1380
cgcagcgcag ccgtggccgt gtttcctctc actcatccca gacacagggt gggggcagcg 1440
tcaccaaaaa gcgcaaactg gagtccactg agagccgcag cagttctca cagcacgcac 1500
gcactagcgg ggcgtgggc cgtggaggag gtggatgagg agggcaagtt tgccggctg 1560
cgcaacaagt ccaatgagga ccagtccatg gcaattggc agatcaagcg ccagaatgga 1620
gatgatccct tgctgactta ccggttccca ccaaagttca ccctgaaggc tggcaggtg 1680
gtgacgatct gggctgcagg agctggggcc acccacagcc cccctaccca cctgggtgtgg 1740
aaggcacaga acacctgggg ctgcgggaac agcctgcgtc cggctctcat caactccact 1800
ggggaagaag tggccatgcg caagctggtg cgctcagtga ctgtgggtga ggacgacgag 1860
gatgaggatg gagatgaccc tccatcac caccacggct cccactgcag cagctcgcccc 1920
ggaccccgct gagtacaacc tgccgtcgcc caccgtgtc tgccggaccc gcccggcagcc 1980
tgccgacaag gcatctgcca gcccggcagg agcccagggtg ggcggaccca tccctctgg 2040
ctcttctgcc tccagtgtca cggtcactcg cagctacccg cgtgtgggggg cagtggggg 2100
tggcagcttc ggggacaatc tggtcaccccg ctccatccctc ctgggcaact ccagcccccc 2160
aaccacagac ccccagaact gcagcatcat gtaatctggg acctgcagg caggggtggg 2220
ggtggaggct tcctgcgtcc tcctcaccc atgcccaccc cctgcctgc acgtcatggg 2280
agggggcttg aagccaaaga aaaataaccc tttgggtttt ttcttctgtt ttttttttc 2340
taagagaagt tattttctac agtggttta tactgaagga aaaacacaag caaaaaaaaaa 2400
aaaaaaagcat ctatctcatc tatctcaatc ctaattctc ctccttcct tttccctgct 2460

33

tccagggaaac tccacatctg ccttaaaacc aaagagggct tcccttagaa gccaaaggaa 2520
 aggggtgctt ttatagaggc tagcttctgc ttttctgccc tgggctgctg cccccacccc 2580
 gggggaccct gtgacatggt gcctgagagg cagggcatag aggcttctcc gccagcctcc 2640
 tctggacgg caggcttcac tgccagggcc agcctccgag agggagagag agagagagag 2700
 gacagcttga gccggggccc tgggtttggc ctgctgtat tccactacac ctggctgagg 2760
 ttcctctgcc tgccccgccc ccagtccca cccctgccc cagccccggg gtgagtccat 2820
 tctccaggt accaagctgc gcttgccttt ctgtatTTTA tttagacaag agatggaaat 2880
 gaggtggag gtggaagaag ggagaagaaa ggtgagttt agctgccttc cctagcttta 2940
 gaccctgggt gggctctgtg cagtcactgg aggttgaagc caagtggggt gctggagga 3000
 gggagagggga ggtcaactgga aaggggagag cctgctggca cccaccgtgg aggaggaagg 3060
 caagaggggg tggaggggtg tggcagtggt tttggcaaac gctaaagagc cttgcctcc 3120
 ccatttccca tctgcacccc ttctctccct cccaaatcaa tacactagtt gttctaaaaa 3180
 aaa 3183

<210> 35
 <211> 207
 <212> DNA
 <213> Homo sapiens

<400> 35

ccaggttgtt ggcgtttcc acagtaactg tgtatgttcc agcatctgtg tcatctgcat 60
 cggtgatggc cagagccgc atcaagccaa tgacgcctgg cacaattcgg ccaggtttct 120
 ccaccacaat ctgccatcc ttctccaga ccacgtcacg ctctttttt aactcgcagc 180
 tcaagtacaa tggctgtctt ttgacca 207

<210> 36
 <211> 253
 <212> DNA
 <213> Homo sapiens

<400> 36

atttatttaca ttttttcatg cactgtcaag ttatctccctt gtccttaac ttctctacag 60
 gatacccttt tctggtttgg ttcatgacaa tctgcaggaa aagagctgcc ttcaaactcc 120
 tttgcttatac ttctccaaaca cttggactc ttgaccgatt ttaccatctc aggtttcaga 180
 gccaggagag agccctgcct catcctgagc tggcatccc catgggtatt ttctgccttt 240
 ctattccctc ttc 253

<210> 37
 <211> 687
 <212> DNA

34

<213> Homo sapiens

<400> 37

tgagccgccc ccgaggattc agcagccctcc cccttgagcc ccctcgcttc ccgacgttcc 60
 gttccccccc gcccccccttc tccccccacc gcccggcccg ccttccgcag gccgggttcc 120
 accgagggaaa aggaatcgta tcgtatgtcc gctatccaga acctccactc tttcgacccc 180
 tttgctgatg caagtaaggg tcatgacctg cttccctgctg gcactgagga ttatatccat 240
 ataagaattc aacagagggaaa cggcaggaag acccttacta ctgtccaagg gatcgctgat 300
 gattacgata aaaagaaaact agtgaaggcg ttaagaaaa agtttgcctg caatggtaact 360
 gtaattgagc atccggaata tggagaagta attcagctac agggtgacca acgcaagaac 420
 atatgccagt tcctcgtaga gattggactg gctaaggacg atcagctgaa ggttcatggg 480
 ttttaagtgc ttgtggctca ctgaagctta agtgaggatt tccttgcata gagtagaatt 540
 tcccttctct cccttgcac aggtttaaaa acctcacagc ttgtataatg taaccatgg 600
 gggtccgctt ttaacttggaa ctgtgtaac tccttgcata aataaactga aaagagccat 660
 gctgtctagt ctgtaaatcc ctcattt 687

<210> 38

<211> 609

<212> DNA

<213> Homo sapiens

<400> 38

ggtgcgggggg cccactgctc tgggctcccc cagggagggaa gcagagtctc gccaagtgc 60
 cctggagggaa tgggagtgaa gcctggcatt ctgaacacat ctctgagggg tgggattaat 120
 aagacggtct ctgtgcctcc tgcctccaga tcctgactgc tgtcatggcg tgccctctgg 180
 agaaggccct ggatgtgatg gtgtccaccc tccacaagta ctggggcaaa gagggtgaca 240
 agttcaagct caacaagtca gaactaaagg agctgctgac cggggagctg cccagcttct 300
 tggggaaaaag gacagatgaa gctgcttcc agaagctgat gagcaacttgc gacagcaaca 360
 gggacaacga ggtggacttc caagagttact gtgtcttcc tgcctgcata gccatgtatgt 420
 gtaacgaatt ctgttgcata ttcccagata agcagccag gaagaaatga aaactctct 480
 gatgtggtttgc gggggctgc cagctggggc cttccctgtc gcccgtggc actttttttt 540
 ttccacccctg gtccttcag acacgtgctt gatgtgagc aagttcaata aagattcttgc 600
 gaagtttta 609

<210> 39

<211> 2539

<212> DNA

<213> Homo sapiens

35

<400> 39

cccccatacat ggttctgctg gagagcaagc attttaccag ggatttaatg gagaagctga 60
aagggagaac cagccgaatt gctggcttg cagtgtcctt gaccaagccc agtccctgcct 120
caggacatct ctccctagtgt acagtgcaca aatgatgggt ttgggtttt aatccaaattcc 180
tatgggccag agtttgctca ctgcagagaa atacagtggaa attcgctggg caatggttt 240
gcttatgaag acttttagttt ccccatctt ctcttgaag atgaaaatga aaccaaagtc 300
atcaagcagt gctatcaaga tcacaacctg agtcagaatg gctcagcacc aacccccc 360
ctatgtgcca tgcagctctt ttcacacatg catgctgtca tcagcactgc cacctgcattg 420
cggcgcagtc catccaaaggc accttcagca tcaacccaga aatcgtctgt gacccctgt 480
ctgattacaa tgggtggagc atgctaaagc ctataaatac aactgggaca tttaaggctg 540
acgacaggggt tgggttgct gccacccggc tggatagtcg ttcccttttc tggaaatgtgg 600
ccccaggggc taaaagcgca gttggcttctt ttgtcacccca gctggctgt gctgaagctt 660
tgcaaaaaggc acctgatgtg accaccctgc cccgcaatgt catgtttgtc ttcttcaag 720
gggaaacttt tgactacatt ggcagctcga ggatggtcta cgatatggag aaggcaagt 780
ttccctgtca gtttagagaat gttgactcat ttgtggagct gggacaggtg gccttaagaa 840
tttcatttgcgac gctttggatg cacacagatc. ctgtttctca gaaaaatggag tctgtacgga 900
accaggtgga ggatctcctg gccacattgg agaagagtgg tgctgggtgc cctgctgtca 960
ttccctcaggag gccaaatcag tcccagcctc tcccaaccatc ttccctgcag cgatttcttc 1020
gagctcgaaa catctctggc gttgttctgg ctgaccactc tgggtgccttc cataacaaat 1080
attaccagag tatttacgac actgctgaga acattaaatgt gagctatccc gaatggctga 1140
gcccctgaaga ggacctgaac tttgttaacag acactgccaa ggccctggca gatgtggcca 1200
cggtgctggg acgtgctctg tatgagcttg caggaggaac caacttcagc gacacagttc 1260
aggctgatcc cccaaacggtt acccgctcgc tctatgggtt tcctgattaa agccaacaac 1320
tcattgttcc agtctatcct cagggcagga cctaaggccc tcttgggtg acgggcctct 1380
tcaacattac atcgctgtct ccagccccac caacaccact tatgtttgtac agtatgcctt 1440
ggcaaatttg actggcacag tggtaacact caccggagag cagtgccagg atccaagtaa 1500
agtccccaaatgt gaaaacaagg atctgtatga gtactcatgg gtccaggggcc ctggcatcc 1560
taatgagacg gaccgactcc cccgggtgtgt gcgttctact gcacgattag ccagggcctt 1620
gtgctcctgc ctggcatcc agtcagtgaa gctctactga atactctaca tggactgaga 1680
ggccctggaa agatatccgt gcccggatat ttctcatcgc cagcaaagag ctggatgtga 1740
tcaccctgac agtgggccttc ggcattcctca tcttctccct catcgacc tactgcattca 1800
atgccaaggc tggatgtccctt ttcatgtgtc cccggggagcc aggagctgtg tcataactgag 1860

36

gaggacccca gcttttcttg ccagctcagc agttcacttc ctagagcata tgtcccactg 1920
 ggacacaacc actaatttgt cactggaacc tccctgggcc tgtctcagat tgggattaac 1980
 ataaaagagt ggaactatcc aaaagagaca gggagaaata aataaattgc ctcccttcct 2040
 ccgcctccct ttcccatcac cccttccca ttcccttcct cttctctact catgccagat 2100
 tttgggatta caaatagaag cttcttgctc ctgtttaact ccctagttac ccaccctaat 2160
 ttgccttca ggacccttct acttttcttcc tccctggccctg tacctctctc tgctcctcac 2220
 ccccacccct gtacccagcc accttcctga ctgggaagga cataaaaggt ttaatgtcag 2280
 ggtcaaaacta cattgagccc ctgaggacag gggcatctct gggctgagcc tactgtctcc 2340
 ttcccaactgt ctttcttca ggccctcaga tggcacatta gggtgggcgt gctgccccgt 2400
 ggtatcccac ctccagccc cagtgtcag ttgtactttt tattaagctg taatatctat 2460
 ttttgggggggtt gtcttttcc tttatcttt tttttttttt atatataatg agtttcttta 2520
 aaatagatta tccccacacg 2539

<210> 40
 <211> 3146
 <212> DNA
 <213> Homo sapiens

 <400> 40

ggagaaggag ctacccccc acctggggga actgaccgtg gctgaggaga cctccagctc 60
 tctgcgcctg tcctggacgg tagcccaagg cccttttgcac tccttcgtgg tccagtacag 120
 ggacacggac gggcagccc gggcagtgcc tgtggccgca gaccagcgca cagtcaccgt 180
 agaggacctg gaggctggca agaaatacaa gtttctgctc tacgggctcc ttggggaaa 240
 ggcgcctgggc cccgtctctg ccctggaaat gacagccccca gaagaggaca caccagcccc 300
 agagtttagcc ccagaggccc ctgagctcc tgaagagccc cgccttaggag tgctgaccgt 360
 gaccgacaca accccagact ccatgcgcct ctctggagc gtggccagg gccccttga 420
 ttccctcgtg gtccagtgatg aggacacgaa cggcggccccc caggccttgc tctggacgg 480
 cgaccagagc aagatcctca tctcaggccc gggcccttca acccccataca ggttcctct 540
 ctatggcetc catgaaggga agcgcctggg qccctctca gctgaggga ccacagggt 600
 ggctcctgtt ggtcagaccc tctggagatc aagggccccggc ctgtcccagc tgtctgtgac 660
 tgacgtgacc accagttcac tgaggctcaa ctgggaggcc ccaccggggg ccttcgactc 720
 cttccctgctc cgctttgggg ttccatcacc aagcactctg gagccgcatac cgcttcact 780
 gctgcagcgc gagctgatgg tgccggggac gggcactcg gcccgtctcc gggacctgctg 840
 ttccgggact ctgtacagcc tgacactgta tgggctgcga ggacccacaca aggccgacag 900
 catccaggga accgccccgca ccctcagccc agttctggag agcccccgtg acctccaatt 960

37

cagtgaaatc agggagacct cagccaaaggta caactggatg ccccccaccaat cccggggcgga 1020
cagcttcaaa gtctcctacc agctggcgga cggaggggag cctcagagtg tgcaggtgga 1080
tggccaggcc cggaccacaga aactccaggg gctgatccca ggcgctcgct atgaggtgac 1140
cgtggctctcg gtcccgaggct ttgaggagag tgagcctctc acaggcttcc tcaccacggt 1200
tcctgacggt cccacacagt tgcgtgcact gaacttgacc gagggattcg ccgtgctgca 1260
ctggaagccc ccccagaatc ctgtggacac ctatgacgac caggtcacag cccctggggc 1320
cccgccctctg caggcggaga ccccaggcag cgcgggtggac tacccttcg atgaccttgt 1380
cctccacacc aactacacccg ccacagtgcg tggcctgcgg ggcccccaacc tcacttcccc 1440
agccagcatac accttcacca cagggctaga ggccccctcg gacttggagg ccaaggaagt 1500
gaccggccgc accggccctgc tcacttggac tgagccccca gtccggcccg caggctacct 1560
gctcagcttc cacacccctg gtggacagaa ccaggagatc ctgctccag gagggatcac 1620
atctcaccag ctcccttggcc tctttccctc cacctcctac aatggcacgg ctccaggcca 1680
tgtggggcca gaggcctctg ccgccccgtgt ccacctctt caccacgggt gggctgcgg 1740
tcccccttccc cagggactgc ggggaggaga tgcagaacgg agccgggtgcc tccaggacca 1800
gcaccatctt cctcaacggc aaccgcgagc ggccccctgaa cgtgttttgc gacatggaga 1860
ctgatgggggg cggctggctg gtgttccagc gccgcattgga tggacagaca gacttctgga 1920
gggactggga ggactatgcc catggtttg ggaacatctc tggagagttc tggctggca 1980
atgaggcccgc gcacagcctg acacaggcag gtgactactc catgcgcgtg gacccgcggg 2040
ctggggacga ggctgtgttc gcccagtagc actccttcca cgtagactcg gctgcggagt 2100
actaccgcctt ccacttggag ggctaccacg gcacccgcagg ggactccatg agtaccaca 2160
gcggcagtgt ctctcttgcc cgtgatcggtt accccaacag ctgcgcaccc tcctgcgtg 2220
tctccttaccg agggggctgg tggtacagga actgcccact acgccaacctt caacgggcgc 2280
tacgggagca cagtggacca tcagggagtg agctggtacc actggaaaggg ctgcgtttc 2340
tcgggtgcctt tcacggaaat gaagctgaga ccaagaaactt ttcgcctcccc agcgggggga 2400
ggctgagctg ctgcccaccc ctctcgacc ccagtagac tggcgagcac tgaggggtcg 2460
ccccggagaga agagccaggg tccttcacca cccagccgtt ggaggaagcc ttctctgcca 2520
gcgatctcgc agcactgtgtt ttaggggggg ggggggggg gttcgtacgg gagcaataaa 2580
ggagaaaactg aggtaccccg ctggcattcg tccctggccca tcactgggttc tggcctgggc 2640
tgtggggcccc catccccccgg ggctgcagcc gcacttggaa aggctgcatac ttgaggatga 2700
caactgcagtg gggcaggggc tgcagggagg gcagggcgac cccggagggc agcagcgtga 2760
aggcctgcag cagtgggtc agcaccacga agagctccag gcgccgcagc ggctcgccca 2820
ggcacacgcg ggcacccgcag ccgaaggcca gagctctgga gttcttgccct ggctccagga 2880

38

agcgatcagg ccagaactca tgtggcctct cccagaccgt ctcatccagg tgggcgcctt 2940
 ggagggttcgg aatgatgact gtgcctcag ggatgtcgta gccagagatg ctgctggcc 3000
 gtgtggtgcg gtggggcaag gctaagggca caacgggccc caggcgcagc acctcggcga 3060
 tggtggcatt gagcaagggc agccgtgcac ggtccttgc 3120
 caccagggcc cagttcgtgg tctagc 3146

<210> 41
 <211> 2898
 <212> DNA
 <213> Homo sapiens

<223> unsure at all n locations
 <400> 41

acagagggac gtggtcactc tctgaaaagt tcaactttag agacaaaatg cagtggacct 60
 ccctcctgtc gctggcaggg ctcttcctcc tctcccaggc ccagtatgaa gatgaccctc 120
 attggtggtt ccactacctc cgcagccagc agtccaccta ctacgatccc tatgaccctt 180
 accgtatga gacctacgag ctttacccct atggggtgga tgaagggcc 240
 acggctctcc atccccctcca gatccccg 280
 acggctctcc atccccctcca gatccccg 300
 acggctctcc atccccctcca gatccccg 360
 cccgcatgaa gtatgtgtac ttccagaaca accagatcac ctccatccag gaaggcgtct 420
 ttgacaatgc cacagggctg ctctggattt ctctccacgg caaccagatc accagtata 480
 aggtggcag gaaggtcttc tccaagctga ggcacctgga gaggctgtac ctggaccaca 540
 acaacctgac ccggatgccc ggtccccctgc ctgcataccct gagagagctc catctcgacc 600
 acaaccagat ctacacgggtc cccaaacaatg ctctggaggg gctggagaac ctacacggct 660
 tgcacccatcca acacaatgag atccaggaag tggcagttc catgaggggc ctccggc 720
 tgcacccatcca acacaatgag atccaggaag tggcagttc catgaggggc ctccggc 780
 ctcttgcata gctgtacatg gacacaaaca atgtctacac cgtcccccgt agctacttcc 840
 gggggggcgcc caagctgctg tatgtgcggc tgcacccatcca cagtctaaacc aacaatggcc 900
 tggcctccaa cacccatcca tccagcagcc tcccttgcata agacccatcc tacaaccaggc 960
 tgcagaagat ccccccagtc aacaccaacc tggagaacct ctacccatcca ggcaatagga 1020
 tcaatgagtt ctccatcagc agcttctgca ccgtggtgga cgtcgtgaac ttctccaa 1080
 tgcagggtctgct ggcctggac gggaaacgaga tcaagcgcag gnccatgcct ggcacgcgc 1140
 ccctctgcct ggccttgcc agcctcatcg agatctgagc agccctggca cgggtactg 1200
 ggcggagagc cccctggca tttggcttga tggtttgggtt tggctttgc tggaaagggtcc 1260
 aggtggacc atgtgacaga agtccacggg caccctctgt agtcttcttt cctgttaggtg 1320
 gggtaggggg gggcgatcag ggacaggcag cttctgctg aggacatagg cagaagctca 1380

ctctttcca gggacagaag tggtggtaga tggaaaggatc cctggatgtt ccaacccat 1440
 aaatctcact gctcttaagt tcttccaaat gatctgaggt catggactt caaaagtggc 1500
 atgggcaata gtatataacc atactttct aacaatccct ggctgtctgt gaggcact 1560
 tgacagctct ccctctgtgc tggctggtc gtgcagttac tctgggctcc catttggc 1620
 ttctcaaaat atacctcttg cccagctgcc tcttctgaaa tccacttcac ccactccact 1680
 ttcctccaca gatgcctctt ctgtgcctta agcagagtca ggagacccca aggcatgtga 1740
 gcatctgccc agcaacctgt ggagacaacc cacactgtgt ctgagggtga aaggacacca 1800
 ggagtcactt ctatacctcc ctaacctcac ccctggaaag ccaccagatt ggaggtcacc 1860
 agcatgatga taatattcat gacctgatgt gggaggagac agccaacctc aggcttagat 1920
 caatgtatag ggctatattt tggcagctgg gtagctctt gaaggtggat aagacttcag 1980
 aagaggaaag gccagacttt ctaccatc agcatctgca atgggccaaa cacacctcaa 2040
 attggctgag ttgagaaagc agccccagta gttccattct tgcccagcac tttctgcatt 2100
 ccaaacagca tcctacctgg ggttttatac cacaaaggta gcggccacat ggttttaaa 2160
 gtatgagaaa cacagttgt cctctccttt tatccaagca ggaagattct atatcctgat 2220
 gtagagaca gactccaggg cagccctggg acttgctagc ccaaagaagg aggatgtgg 2280
 taatctgttt cacctgggtt gtcctaaggc catagttaaa aagtaccagc tctggctggg 2340
 gtccgtgaag cccaggccag gcagccaaat ctggcctgt gctgggcata caaccctctg 2400
 cttcacatc tctgagctat atcctcattt gtgaaggtgg ctggctttt atagttggc 2460
 tggggagcac ttaattcttc ccatttcaaa aggtaatgtt gcctggggct taacccacct 2520
 gccccttggg caagggttggg acaaagccat ctgggcagtc aggggcaagg actgttggag 2580
 gagagttagc ccaagtatag gctctgccc gatgccatca catccctgat actgtgtatg 2640
 ctggaaagca cttccctga gaaggaaaga gggatcttt ggacttaggtt ctggctcca 2700
 gacctggaat ccacaaaagc caaaccagct catttcaaca aaggagctcc gatgtgaggg 2760
 gcaaggctgc cccctgcccc agggctcttc agaaagcatc tgcattgtgaa caccatcatg 2820
 ctttataaaa ggatccttat tacagggaaaa gcatgagtggtt tggctaacct gaccaataaa 2880
 gttattttat gattgcaa

2898

<210> 42
 <211> 854
 <212> DNA
 <213> Homo sapiens

 <223> unsure at all n locations
 <400> 42

ttccggcacag cgnngggata caactctggaa gtcctctgag agagccacca aggaggagca 60

40

ggggagcgac ggccggggca gaagttgaga ccacccagca gaggagctag gccagtccat 120
 ctgcatttgt cacccaaagaa ctcttaccat gaagaccctc ctactgttgg cagtgtatcat 180
 gatcttggc ctactgcagg cccatggaa ttggtgaat ttccacagaa tgatcaagtt 240
 gacgacagga aaggaagccg cactcagtta tggcttctac ggctgccact gtggcgtggg 300
 tggcagagga tcccccaagg atgcaacgga tcgctgctgt gtcactcatg actgttgcta 360
 caaacgtctg gagaaacgtg ggatgtggc accaaatttc tgagctacaa gtttaggcaa 420
 ctcggggagc agaattcacct gtgcaaaaca ggactcctgc agaagtcaac tgggtgagtg 480
 tgataaggct gctgccacct gttttgctag aaacaagacg acctacaata aaaagtacca 540
 gtactattcc aataaacact gcagagggag caccctcgt tgctgagtcc cctttccct 600
 ggaaaccttc cacccagtgc tgaatttccc tctctcatac cctccctccc taccctaacc 660
 aagttccttg gccatgcaga aagcatccct cacccatcct agaggccagg caggagccct 720
 tctataccca cccagaatga gacatccagc agatttccag ctttctactg ctccctcca 780
 cctcaactcc gtgcttaacc aaagaagctg tactccgggg ggtctttct gaataaagca 840
 attagcaaat catg 854

<210> 43
 <211> 471
 <212> DNA
 <213> Homo sapiens

<400> 43

caataccatg aagaggaggc tcaggcagct cttaccacat gatacaagag ccggctggtg 60
 gaagagtggg gaccagaaag agaatttgct gaagaggaga agaaaaaaa aaacaccaaa 120
 aaaaaaaaaata aaaaaatcca cacacacaaa aaaacctgcg cgtgaggggg gagaaaaagc 180
 agggcccttt aaaaaggcaa tcacaacaac tttgctgcc agggatgccc ttgctttggc 240
 tgagaggatt tctgttggca agttgctgga ttatagttag gagttccccc accccaggat 300
 ccgagggggca cagcgcggcc cccgactgtc cgtccctgtgc gctggccgccc ctcccaaagg 360
 atgtacccaa ctctcagcca gagatggtgg aggcgtcaa gaagcacatt taaaacatgc 420
 tgcacttgaa gaagagaccc gatgtcaccc agccggtacc caaggcggcg c 471

<210> 44
 <211> 1411
 <212> DNA
 <213> Homo sapiens

<400> 44

gccactgctc tgagaatttg tgagcagccc ctaacaggct gttacttcac tacaactgac 60
 gatatgatca tcttaattta cttatccctc ttgctatggg aagacactca aggatgggg 120

41

ttcaaggatg gaattttca taactccata tggcttgaac gagcagccgg tgtgtaccac 180
 agagaagcac ggtctggcaa atacaagctc acctacggca gaagctaagg cggtgtgtga 240
 atttgaaggc ggccatctcg caacttacaa gcagctagag gcagccagaa aaattggatt 300
 tcatgtctgt gctgctggat ggatggctaa gggcagagtt ggatccccca ttgtgaagcc 360
 agggcccaac tggggatttg gaaaaactgg cattattgtat tatggaatcc gtctcaatag 420
 gagtgaaaga tgggatgcct attgctacaa cccacacgca aaggagtgtg gtggcgtctt 480
 tacagatcca aagcaaattt ttaaatctcc aggttccca aatgagtagc aagataacca 540
 aatctgctac tggcacatta gactcaagta tggtcagcgt attcacctga gttttttaga 600
 ttttgcctt gaagatgacc caggttgcct ggctgattat gttgaaatat atgacagtta 660
 cgatgatgtc catggctttg tgggaagata ctgtggagat gagcttccag atgacatcat 720
 cagtacagga aatgtcatga ctttgcgtt tctaagtgtat gtttcagtga cagctggagg 780
 tttccaaatc aaatatgttg caatggatcc tggatccaa tccagtcaag gaaaaataac 840
 aagtactact tctactggaa ataaaaactt ttttagctgga agatttagcc acttataaaa 900
 aaaaaaaaaag gatgatcaaa acacacagtg tttatgttgg aatcttttgg aactcctttg 960
 atctcactgt tatttataac atttattttat tatttttcta aatgtgaaag caatacataa 1020
 ttttagggaaa attggaaaat ataggaaact taaaacgaga aatgaaacc tctcataatc 1080
 ccactgcata gaaataacaa gcgttaacat tttcatattt ttttcttca gtcattttc 1140
 tatttgcgtt atatgtatat atgtacctat atgtatgtc atttgcgtt ttggatcc 1200
 gctctatgtt cagttttgtt ttatactttt taaatcttga actttataaa cattttctgtt 1260
 aatcattgtat tatttctacaa aaacatgatt taaaacagct gtaaaaatatt ctatgatatg 1320
 aatgttttat gcatttatttta agcctgtctc tattgttggaa atttcagggtc attttcataaa 1380
 atattgttgc aataaatatc cttgaacacaca c 1411

<210> 45
 <211> 1877
 <212> DNA
 <213> Homo sapiens

<400> 45

gttcttgcct agtgagcaga tccaggggt tggatctcc gtgattaacc tggaggctag 60
 aactggcttc ttgtccaaacc cttagggcctg gggccgcattt gacagtgtca tcacaggccc 120
 caacggggcc tggatgttgcct gccttctgtt atgaccagtc ccctgatgcc tactctgcct 180
 atgtcttggc aagcctggct ggggaggaac tgcaagcagt gggagtcttc tcctaaattc 240
 aacccaaatg caattggcgt ccctcagccc tatctcaaca agctcaacta ccgtcggacg 300
 gaccatgagg atccacgggt taaaaagaca gcttccaga ttagcatggc ccaagccaag 360

42

gcccaactca gctgaggaga gcaatggcc catctatgcc tttgagaacc tccgggcatt 420
 tgaagaggca ccacccagtg cagcccaatt ccgggttctac cagattgagg gggatcgata 480
 tgactacaac acagtcctt tcaacgaaga tgaccctatg agctggactg aagactatct 540
 ggcatggtgg ccaaagccga tggaaattcag ggcctgctat atcaagggtga agattgtggg 600
 gccactggaa gtgaatgtgc gatcccgaa catggggggc actcatggc ggacagtggg 660
 gaagctgtat ggaatccgag atgtgaggag cactcgggac agggaccagc ccaatgtctc 720
 agctgcctgt ctggagttca agtgcagtgg gatgctctat gatcaggacc gtgtggaccg 780
 caccctggtg aaggtcatcc cccagggcag ctgcccgtcga gccagtgtga accccatgt 840
 gcatgagttac ctggtaacc acttgcact tgcagtcac aacgacacca gtgagttac 900
 catgctggca cccttggacc cactggcca caactatggc atctacactg tcactgacca 960
 ggaccctcgc acggccaagg agatcgccgt tggggccgtg ctgtatggc acatccgatg 1020
 gctccctccag aatcatgaag agcaatgtgg gatcgccct caccttcaac tgttagaga 1080
 ggcaagtagg cccgcagagt gccttccagt aatccaaag cacccagcc cagtcctctg 1140
 ctgcaggcac tgcaccaagg agagtgcctt cggggggca gcagcgagcg agcaggggtg 1200
 gccagcgcca gagtggagtg gtggcccttc tggatttcc tagagttgtt caacagcccc 1260
 tgcataacta agttttgtgg tacttaccc tttctgtccc tcatttcatg tgacagccat 1320
 tgcataactg atgcacaaac tgcacccgg taaatccaaatg cacttctgtt ttctgttatt 1380
 tgcataactg tttcttcatg cttttactt tttgtccca tgcatactgat tggcacgtgg 1440
 ccccaataat ggcacaataa agcccccttg taaaactgtt ctttaatgaa aacacaagaa 1500
 atggccact ggtaaaactc tgcacccgttca actgtacccat attaatgccc attaatgca 1560
 atataacttcc tttttttttt gcatggttt gcccacccctt gcaatagtga taatctgtat 1620
 ctgaagatca aataaccaat ataaagcata tttttggcc tgcacccaca ggacataggc 1680
 aaggcccttga tcatagttca tacatataaa tttttttttt gaaatggaa ataaagaaat aaaacacaat 1740
 acttttactt gaaatgtaaa taacttattt tttttttttt gaaatggaa attctagtgc 1800
 acattccaaag ttaagctatt aaatataaggc tttttttttt gcaatagtgc tttttttttt gaaatggaaa 1860
 agaacatctc ctggat 1877

<210> 46
 <211> 167
 <212> DNA
 <213> Homo sapiens

<400> 46

atcaaaaaaca tcaactccctc tccctccctt acaatggaaa gagagaaggg agactctatt 60
 taagattttttt aacaccaatg atcatctgaa tcccgccatg agaatgcaga cttttcagac 120

43

tgaccccaga aattctggcc cagccaatct agaggcaagc ctggcca

167

<210> 47
 <211> 1689
 <212> DNA
 <213> Homo sapiens

<400> 47

ccgcctccg ccacctttct tgggtggctc tccgcctcgt cctccctccg agggccgttg 60
 gtacattcct agtgactcca agcgcttaaa aggggcccgg gaggatgaac cccacagatc 120
 tgaacctgat ttgtgtgtgc accgcgtctc cagcgatccc ggatccactg cgctgccagg 180
 gcgcctgggg gtgggtctct tgctgtctct ggcacgacat cttacgttt cggcactcta 240
 atgctgggtt tggctgtgtg tggctgtta gggctctagc gggctgttag gctccctcgc 300
 cccagctcc ttggctcgct cagctccctcc accgcagccc agcagtgaga cgcgcgcgca 360
 gccagctccc cacgagatgg aacagaccga agtgctgaag ccacggaccc tggctgatct 420
 gatccgcattt ctgcaccaggc tctttgccgg cgatgaggc aatgttagagg aggtgcaggc 480
 catcatggaa gcttacgaga gcgacccac cgagtggca atgtacgcca agttcgacca 540
 gtacaggtat acccgaaatc ttgtggatca aggaaatggaa aaatttaatc tgatgattct 600
 ctgtgggtt gaaggacatg gcagcagtat tcatgatcat accaactccc actgctttct 660
 gaagatgcta cagggaaatc taaaggagac attatttgc tggcctgaca aaaaatccaa 720
 tgagatggtc aagaagtctg aaagagtttt gagggaaaac cagtgtgcct acatcaatga 780
 ttccattggc ttacatcgag tagagaacat cagccatacg gaacctgctg tgagccttca 840
 cttgtacagt ccacctttt atacatgcca tgcctttgat caaagaacag gacataaaaa 900
 caaagtaca atgacattcc atagtaaatt tggaatcaga actccaaatg caacttcggg 960
 ctgcgtggag aacaactaag gggcacaaaa ccctctgagg ttttacttta aggttcgctg 1020
 tatgtttgcc ttggacaaaa aggctaccta ccacgtgcta tccagtaata tacttaata 1080
 agccaatact tagatctact gtaaggcaga tgcttaattt aaggcattaa gtaagcaaat 1140
 agtgcctca gctactgcag aagaaaaagtc ccactgagga aaagaaaagtc ttgtgattt 1200
 taaaggcaag tttcaagtg ctctcatagt tctatcctct aattccattt aatccatact 1260
 aggagcgtca gtgagggttt tcatagttt tggaaatact ttggctctg aactgttaatt 1320
 agcaagaagt aaaaacagaa acgtcaaacg tcaaattttt gctttgttac ctggaggact 1380
 aatgttagat gtcttttagta tactttgtat gttcttaata ttggaaagata attttgtaa 1440
 tctgttagatt ttatttttc agtcttacct tacaatttc ttttctatga ataatagagg 1500
 aacttacggc actctgccat ttgttaatga aaggaagtgc agaggattt gaaaagtaca 1560
 tcatccccag accacaacaa accaaaacat aaactcatgt ctgtgtccca tggctatagt 1620

44

caaagat~~ttt~~ gtactgctaa aattac~~ccaa~~ taat~~tt~~aaat aaagtggatt tgaacaca~~aa~~ 1680

aaaaaaaaaa

1689

<210> 48
<211> 184
<212> DNA
<213> Homo sapiens

<400> 48

agaaaacaat gaagaatcga atgaagatga agactctgag gctgagaata ccacactt~~tc~~ 60

tgctacaaca ctgggctatg gagaggacgc cacgcctggc acagggtata cagggtagc 120

tgcaatccag ctccccaa~~ga~~ aggctgggg~~a~~ tataacaa~~ac~~ aaagctacaa aagagaagga 180

aagt 184

<210> 49
<211> 259
<212> DNA
<213> Homo sapiens

<400> 49

cctggccccc~~g~~ tgggt~~c~~ctcc tggcctgacg ggtcctgcag gtgaacctgg acgagaggga 60agccccgg~~t~~ ctgatggccc ccctggcaga gatggcgctg ctggagtc~~a~~ ggg~~t~~gatcgt 120gg~~t~~gagactg gtgctgtgg~~g~~ agctcctgg~~a~~ gccc~~c~~ttgggc cccctggctc ccctggcccc 180

gctggtccaa ctggcaagca aggagacaga ggagaagctg gtgcacaagg ccccatggga 240

ccctcaggac cagctggag 259

<210> 50
<211> 245
<212> DNA
<213> Homo sapiens

<400> 50

gagagaagg~~g~~ ccacccaggt ctcattggac tgattggcc cccgggtgag cagggagaga 60agggagatcg gggacttcct gggcctcagg gctccc~~t~~gg gcagaagg~~g~~ gagatgg~~g~~ta 120tcccagg~~g~~ac atccggcccc attgg~~t~~cctg gaggtcccc cggcctcccc ggac~~c~~tcgt~~g~~ 180gccccaaagg agccaaagga gccacaggcc caggcggacc caagggagag aagg~~g~~tg~~t~~gc 240

agggc 245

<210> 51
<211> 515
<212> DNA
<213> Homo sapiens

<400> 51

45

cttgcagaga aagagtcttt tgtgcagcac cctttaaagg gtgactcgac ccacttgtgt 60
 tctctctcct ggtgcagagt tgcaagcaag tttatcagag tatcgccatg aagttcgcc 120
 cctgccttct gctggtgacc ttgtccgtcc tggggacttt gggtcaggcc ccgaggcaaa 180
 agcaaggaag cactggggag gaattccatt tccagactgg agggagagat tcctgcacta 240
 tgcgtcccgag cagcttgggg caaggtgctg gagaagtctg gcttcgcgtc gactgccgca 300
 acacagacca gacctactgg tgtgagtaca gggggcagcc cagcatgtgc caggcttgc 360
 ctgctgaccc caaatcttac tggaatcaag ccctgcagga gctgaggcgc ctaccatg 420
 cgtgccaggg ggccccgggtg cttaggcat ccgtgtgcag ggaggctgga cccaggccc 480
 atatgcagca ggtgacttcc agcctcaagg gcagc 515

<210> 52
 <211> 281
 <212> DNA
 <213> Homo sapiens

<400> 52

gccccggggcc ctggacgatg tggagaacct cgccaaattc cacgtggaca ggaaccagct 60
 gtccagctac ccctcagctg ccctgagcaa gctacgggtg gtggaggagc tgaagctgtc 120
 ccacaacccc ctgaaaagca tcccgacaa tgccttccag tcctttggca gataacctgga 180
 gaccctctgg ctggacaaca ccaacctgga gaagtctca gatggtgct tcctgggtgt 240
 aaccacgctg aaacacgtcc atttggagaa caaccgcttg a 281

<210> 53
 <211> 252
 <212> DNA
 <213> Homo sapiens

<400> 53

gggacagatc ccaggggtgcc cagggagtct ccaagtgcct cactcctccc gccgcaaaca 60
 tgacagagaa ctccgacaaa gttccattt ccctgggtgg acctgatgac gtggaaattct 120
 gcagcccccc ggctgtacgct acgctgacgg tgaagccctc cagccccggc cggctgctca 180
 aggtgggagc ctgtggtcctc atttggggag ctgtgtctgct gctctttggg gccatcgggg 240
 ctttctactt aa 252

<210> 54
 <211> 2723
 <212> DNA
 <213> Homo sapiens

<400> 54

gacatagctt ttctcattca ccctcccact tggggctaat gcacagacat gaacatctat 60

46

tgaggaaaac cacaaaaaac ttcaaaacag ctacaacggg aaaaagagag ttttgtccca 120
cagtcagcag gccactagtt tattaacttc cagtcacctt gattttgct aaaaatgaaga 180
ctctgcagtc tacacttctc ctgttactgc ttgtgcctct gataaagccc aggcaccacc 240
aaccaggcag gactcacgca ttatctatga ttatggaaca gataatttg aagaatccat 300
attagccaa gattatgagg ataaataacct ggtggaaaa aatattaagg aaaaagaaac 360
tgtgataata cccaatgaga aaagtcttca attacaaaaa gatgaggcaa taacaccatt 420
acctcccaag aaagaaaaatg atgaaatgcc cacgtgtctg ctgtgtgtt gtttaagtgg 480
ctctgtatac tgtgaagaag ttgacattga tgctgtacca cccttaccaa aggaatcagc 540
ctatcttac gcacgattca acaaattaa aaagctgact gccaaagatt ttgcagacat 600
acctaactta agaagactcg atttacagg aaatttgata gaagatatag aagatggtag 660
ttttcaaaa ctttctctgt tagaagaact ttcaacttgct gaaaatcaac tactaaaact 720
tccagttctt cctcccaagc tcactttatt taatgcaaaa tacaacaaaa tcaagagtag 780
gggaatcaaa gcaaattgcat tcaaaaaact gaataacctc accttcctct acttggacca 840
taatgccctg gaatccgtgc ctcttaattt accagaaaagt ctacgtgtaa ttcatcttca 900
gttcaacaac atagcttcaa ttacagatga cacattctgc aaggctaattg acaccagtta 960
catccgggac cgcattgaag agatacgcct ggagggcaat ccaatcgcc tgggaaagca 1020
tccaaacagt tttatggct taaaaagatt accgataggg tcatactttt aacctcttatt 1080
ggtacaacat ataaatgaaa gtacacccat actaatagtc tgtctcaaca atgagtaaag 1140
gaacttaagt attggttaa tattaacctt gtatctcatt ttgaaggaat ttaatatttt 1200
aagcaaggat gttcaaaatc ttacatataa taagaaaaa gtaagactga atgtctacgt 1260
tcgaaacaaa gtaatatgaa aatattttaa cagcattaca aaatccatgt ttatactaga 1320
ctaccattta aaaatcatgt ttttatataa atgcccaat ttgagatgca ttattccat 1380
tactaatgat gtaagtacga ggataaatcc aagaaacttt caactcttg cctttcctgg 1440
cctttactgg atccccaaag catttaaatggt acatgttcca aaaactttga aaagctaaat 1500
gtttcccatg atcgctcatt cttctttat gattcatacg ttattccctta taaagtaaga 1560
actttttt cctccatca aggcagctat ttattttaaat tttcactta gtctgagaaa 1620
tagcagatag tctcatattt aggaaaaactt tccaaataaa ataaatgtta ttctctgata 1680
aagagctaat acagaaatgt tcaagttatt ttactttctg gtaatgtctt cagtaaaaata 1740
ttttctttat ctaaatatttta acattctaag tctacaaaaa aaagttttaa actcaaggcag 1800
gccaacca atatgcttat aagaaataat gaaaagttca tccatttctg ataaagttct 1860
ctatggcaaa gtctttcaaa tacgagataa ctgcaaaata ttccctttt atactacaga 1920
aatgagaatc tcatcaataa attagttcaa gcataagatg aaaacagaat attctgtgg 1980

47

gccagtgcac actacccatcc caccataca catccatgtt cactgtaca aactgaat 2040
 tcacaataaa gcttctgagt aacactttct gattactcat gataaactga catggctaac 2100
 tgcaagaatt aaatcttcta tctgagagta ataatttatg atgactcagt ggtgccagag 2160
 taaagttct aaaataacat tcctctcaact tgtacccac taaaagtatt agtctacaca 2220
 ttacattgaa gttaaacaca aaattatcag tgtttagaa acatgagtcc ggactgtgta 2280
 agtaaaaagta caaacattat ttccaccata aagtatgtat taaaatcaag ttgtctctgt 2340
 gtacagaata catacttatt cccatttttta agcatttgct tctgtttcc ctacctagaa 2400
 tgtcagatgt ttttcagttt tctcccttatt tgtaaagttt gaccaatcaaga taacattttt 2460
 cattaaagca tctgagatct aagaacacaa ttattattct aacaatgatt attagctcat 2520
 tcacttattt tgataactaa tgatcacagc tattatacta ctttctcgat attttgtgtg 2580
 catgcctcat ttccctgact taaacctcac tgagagcgca aatgcagct ttatactttt 2640
 tactttcaat tgccctagcac aatagtgagt acatggaaat tgaatatata ataaatattg 2700
 caaaaataaaaa tccatctaaa tag 2723

<210> 55
 <211> 310
 <212> DNA
 <213> Homo sapiens

<400> 55

ggcgcggcc gcccgtgtc cccccagccc cggccccagg cgtcccgcc atggccgccc 60
 caatgtctt gtcagcctc ggcctctgg ctggtctgtc gccggcgctg gccgcctgccc 120
 cccagaactg ccactgccac agcgacctgc agcacgtcat ctgcgacaag gtggggctgc 180
 agaagatccc caaggtgtca gagaagacca agctgctcaa cctacagcgc aacaacttcc 240
 cgggtgtggc tgccatttcg ttccggccca tgccgaacct cgtgtcatttgc caccgtcagc 300
 actgcccagat 310

<210> 56
 <211> 274
 <212> DNA
 <213> Homo sapiens

<400> 56

atttatgaaa tcataaaacc tgcaacagcc aactcgaaat tccccgtgac cagtcttttg 60
 gacaccaggg acagcaatga gcctgactct cctgcatttc ctttgtctga ggcataagacc 120
 actgactgct tatggaaaag aacagataat gatatccgtc tccgtctcc acccaccact 180
 caatgtact ttctgccatg aacataacca gcccacacata aactgtctgc agaaaaggaa 240
 gttccatcct ataagcttgg caggaggata aaga 274

<210> 57
 <211> 153
 <212> DNA
 <213> Homo sapiens

<400> 57

aattttaaga tttaactta cacaaaaagt ccacttacaa gcatttatct catttacatg 60
 tattcacctt ttccatttct taatagttt tctagattac ttctgaaaac tgagatatta 120
 cacaaaacta atcattattt aaagttatcc 153

<210> 58
 <211> 225
 <212> DNA
 <213> Homo sapiens

<400> 58

tgatggtaag ttgttcagg cataaaatcc gaaataaattt atgaggctcc atgatatgct 60
 atattggttt taccttcaga agaatatcc gtttactca gtttttcaa agctacgctg 120
 tcccccaaaa aacgaaacaa aacaaaaaaaaa caacccccc aagagttgat ggctactcat 180
 ttgatctgcc tcctctgctg aatcaattttt gaaatccccc tttttt 225

<210> 59
 <211> 448
 <212> DNA
 <213> Homo sapiens

<400> 59

ggaagcgtcc aaagagggac ggctgtcagc ccgggttga ctgagaaccc accagctcat 60
 cccagacacc tcatacgcaac ctatccatac aaagggggaa agaaacacct gagcagaatg 120
 gaatcattat tttttccca aggagaaaaac cggggtaaaag ggagggaaagc aattcaattt 180
 gaagtccctg tgaatgggct ttcagaaggc aatccaaagaa atccactcag agaggacttg 240
 ggggtgaaact tgggtccctgt ggtttctga tttaatgg aagcaggctt tgcacacgct 300
 gttggcaaat gtcaggacca ggttaagtga ctggcagaaa aacttccagg tggaaacaagc 360
 aaccaggctt ctgctgcaag ctggaaagga gctggagcg ggagaaagct aacttgaaca 420
 tgacctgttg catttggcaa gttcttagc 448

<210> 60
 <211> 59
 <212> DNA
 <213> Homo sapiens

<400> 60

atgacattgg ttgcctcagc cctgaaaagc tttgtctctg cattcttagt tttctttgt 59

<210> 61
 <211> 321
 <212> DNA
 <213> Homo sapiens

<223> unsure at all n locations
 <400> 61

attaattgcc agtagttgta aggaggagtc agcatctagt gttactccct nnnnnnnnnnn 60
 nnnnnnnnnnn nnnntccagg tactggctaa tggagctact gccacctcta aacccttcca 120
 gccacttaggc tgtgtcccac agtcagtgtc acccagtgaa caggcattac ccccacatct 180
 ggaaccagcc tggccccaag ggctacggca taacttagta ccaggttagag ttggcccccac 240
 agagtacctt tccccagata tgcaacgcca gcgaaagacc aagcgaaaaa ccaaagagca 300
 gctggctatc cttaaatcct t 321

<210> 62
 <211> 252
 <212> DNA
 <213> Homo sapiens

<400> 62

tttccctaat atttaaatta ttccttataa accagtagaa aagcttaac aacataacag 60
 aaaaatggga aaagactatg aatagacggg acccagaaaa gcacatacaa ataagtggct 120
 attttactac acctttactt tggaaaactt caaacctgta ctaaaataga atagggcagt 180
 gaacctccct gcctgcaccc atcactcagc gtcaacattt atcaactcat gggcaatctt 240
 gtttatcta tt 252

<210> 63
 <211> 218
 <212> DNA
 <213> Homo sapiens

<400> 63

cacaagttaa aacttcccat gtataaaaac acttacattt taaaacatca ctgccaactg 60
 tgtgctcatg tgggagtaca gatgtgtata tacagacatg tacatttta aagacttggt 120
 tgtctctgca gtgaagacaa tatgtttat ttttattcc atataacttct ctgtatttc 180
 tatatttgct tcaataagct ggtgtaactt ttaatttt 218

<210> 64
 <211> 235
 <212> DNA
 <213> Homo sapiens

<400> 64

gatcaaatacg gaaaggtaaa gatgaaatgc ttttccgtt tcttgatttt tatctaccag 60

50

caataatatg aggcacactc gtaaaatgtaaa gggtttgcatt atatttacaa ttaaaactcta 120
 gaaaagcata attctgagct aaatattctg cctaaagaat ctctttcaca taatccttcc 180
 tggtcacttg ctcccttgcac tcacaatgg tttcttaatt cctatgcttt ttatc 235

<210> 65
 <211> 239
 <212> DNA
 <213> Homo sapiens

<400> 65

tgccgcgttg ttgagccctt aaaataccac ctccctcatgt gtaaaattgac acaatcacta 60
 atctggtaat taaaacaatt gagatagcaa aagtgtttaa cagacttagga taattttttt 120
 ttcatatgg cccaaatttt tgtaaaccct gtctgtcaa ataagtgtat aatattgtat 180
 tattaaattta ttttactttt ctataccatt tcaaaacaca ttacactaag ggggaacca 239

<210> 66
 <211> 243
 <212> DNA
 <213> Homo sapiens

<400> 66

ggaaactcca ggctcctggg tttccctgg gcggggaaag agaagactga aacatctgtg 60
 tgacattcag attttcaga ggtctgccca agggctctggg ttttattttt cttgaatata 120
 agttctgaca ggaaaggca ccaggttgcg gggtcattga aaacaaagtt gacagtttag 180
 attagcaggc actcaccatg gtccctcccc ctccctcagc atgaaaacca gcaggagaaa 240
 ttc 243

<210> 67
 <211> 250
 <212> DNA
 <213> Homo sapiens

<400> 67

gtctgtgtac catcttaccc ggaatagaga ttgtgttaaa ttaacagatc atctgactga 60
 gaggtttttt tcccccaaaa cagaagcaaa taaacattat ttgttcctt tggtataact 120
 ttcattgaac agttatatacg tgctttggaa gtatcaagtc ctgtgctaaa taaatgctgg 180
 agatacaaaa gccctgacc tcagaatgtc atagtttgg ggtaagaaaa aattcattct 240
 gtgccccagg 250

<210> 68
 <211> 213
 <212> DNA
 <213> Homo sapiens

<400> 68

caggtgtgaa ccactgcacc tggcccaaaa tctcttgatt gatacagtcc tctttat 60
tcaagatcaa gttatgatac ctttaccaac agtcatacat tctttggaa ctttgcacaa 120
tagtcatatg ttcttttaga actttacact tctattctt attgccctgt attataattg 180
cttgtatgcc tgactcctct acatgactgt atg 213

<210> 69

<211> 198

<212> DNA

<213> Homo sapiens

<400> 69

cataaaccta ctttatcate ctctcctaaa gggaaaagag aagatttagc tagaataatt 60
attaacagaa gatgtggaga tacagaagaa actagaaaat atctcacaat caatacatct 120
ttcaagcagt caatcatttg tcactcatat tgcttttta aacccagctt tacatggaag 180
gaataaatgg aactccag 198

<210> 70

<211> 393

<212> DNA

<213> Homo sapiens

<400> 70

aaaaaaaaagga aaaaaaaaaat tgccttaagt catatagatt gtaccagcag ctctcacagt 60
gtggactttg gacttctagg agtccccagg aaccttttag gggatgccta cgaggaggtc 120
caaactgttt tcataagaac gctaagggtgc tatgtgcctt tttaactcat tcttcacga 180
gtgttcagtg gagttttcca gaggctctgt gacatggtga catcaactctg ataatttagta 240
gaatgtgtgt gtgtgtactt ttgtttctta gaatattgtt aattgataga tttagggtat 300
aaatatatgt gttttcagag attaactcag tttgctgcca gtgcttctac tgtgctctta 360
ctggctattt tcatttatac ctgctgctga gtc 393

<210> 71

<211> 216

<212> DNA

<213> Homo sapiens

<400> 71

ctctacttgt atgaccctag gaatagattt gaatactgca gaggaccaaa gctgaggcat 60
gctaaacagc tgcttggagg tggaaagcaag ttcaagtcacc tactcagctt cctctctcca 120
ccacccagtt cctccctcag tatacattt ttttttctt ctgctttca ttaacctaac 180
tcatctcattc agtacaacca ttttcttatt ctctaa 216

<210> 72
 <211> 166
 <212> DNA
 <213> Homo sapiens

<400> 72

caaatat~~ttt~~ta acagaactaa t~~gg~~aaactatt ttagtatgct ttccc~~ttt~~ggg ctggagt~~ttt~~ta 60
 ggcttaagact ttat~~ttt~~aaat acaggat~~gg~~ga t~~gg~~tg~~ttt~~tg actgaagatg cctccaactt 120
 ttgctcttct gtttttatt tgatgtgctc aagcttctaa t~~ttt~~ccct 166

<210> 73
 <211> 240
 <212> DNA
 <213> Homo sapiens

<400> 73

tgataggcag ctaaaactgt tatgccact gtgctcaatt tgaagcagaa ttcagt~~ttt~~aaa 60
 aattat~~ttt~~ttt ccacatt~~gg~~aa acacttt~~gg~~ca gacacaaata tctatgaaaa gatgctttgt 120
 cagccactgt gcctttttt ctgtgaagac tcaacggatg t~~gt~~gtg~~ttt~~tg tatg~~ttt~~gtt 180
 aacagttaca tatg~~ttt~~gt~~ttt~~ta tgagtgtata tatatatctg t~~gt~~gtg~~ttt~~gt~~ttt~~ta tctctaacgt 240

<210> 74
 <211> 291
 <212> DNA
 <213> Homo sapiens

<400> 74

tggaccccca gctgaggagt cctgctcaag acacggcac tggatctgag aaacttccca 60
 ggggacccgca ttccagagtc agtgactctg tgaagcaccc acatctaccc tttgccacgt 120
 tcccacgggc ttgggggaaa gatgggtggg accaaggcct ggg~~ttt~~ctc cttcc~~ttt~~ggc 180
 ctggaagtca catctgtgtt ggggagacag acgtgctca cccagtcagt aagaagagtc 240
 cagcctggga agaagaaccc cagcatctt gccaagcctg ccgacacccct g 291

<210> 75
 <211> 283
 <212> DNA
 <213> Homo sapiens

<400> 75

ctccggc~~gg~~gc ctccgggaga ggagccgcac cccggccggcc cggcccc~~gg~~gc cccatggacc 60
 tccgagcagg ggactgcgtg ggggatgtta gcgtgcctgt gcacggtgct ctggcaccc 120
 cctgcagtgc cagctctcaa tcgcacaggg gacccagg~~gg~~gc ctggccctc catccagaaa 180
 acctatgacc tcacccgcta cctggagcac caactccgca gcttggctgg gacctatctg 240

aactacacctgg gcccccccttt caacgagcca gacttcaacc ctc 283

<210> 76
<211> 139
<212> DNA
<213> Homo sapiens

<400> 76

ccttcgtgaa gtcgccaaac ctctctgagc cccagtcatt gctagtaaga cctgcctttg 60
agttggtagt atgttcaagt tagataacaa aatgtttata cccattagaa cagagaataa 120
atagaactac atttcttgc 139

<210> 77
<211> 669
<212> DNA
<213> Homo sapiens

<400> 77

ctggctggag cagcgagtct gtcgatccca ggccagagac aaggcagaca aaggttcatt 60
tgtaaagaag ctccctccag cacctcctct cttctccctt tgcccaaact cacccagtga 120
gtgtgagcat ttaagaagca tccctgtcca agaccaaaag gaaagaagaa aaagggccaa 180
aagccaaaat gaaactgatg gtacttgttt tcaccattgg ggctaacttt gctgctagga 240
gttcaagcca tgcctgcaaa tgcctctct tgctacagaa agatactaaa agatcacaac 300
tgtcacaacc ttccggaagg agtagctgac ctgacacaga ttgatgtcaa tgtccaggat 360
catttctggg atgggaaggg atgtgagatg atctgtact gcaacttcag cgaattgctc 420
tgctgccaa aagacgtttt ctttggacca aagatctctt tcgtgattcc ttgcaacaat 480
caatgagaat cttcatgtat tctggagaac accattcctg atttcccaca aactgcacta 540
catcagtata actgcatttc tagttctat atagtgcatt agagcataga ttctataaaat 600
tcttacttgt ctaagacaag taaatctgtg ttaaacaagt agtaataaaa gttaattcaa 660
tctaaaaaa 669

<210> 78
<211> 486
<212> DNA
<213> Homo sapiens

<400> 78

ggacgccatc tctgaggccc aaggccacag tgaaatcaca gaagcaacac agctggaaaa 60
ggactcgatg gaagagctgg gaaaagccaa acccaccacc cgacccacag ccaaacctac 120
ccagcctgga cccaggccccg gagggaatga ggaagcaaag aagaaggcct gggAACATTG 180
ttggaaaccc ttccaggccc tgcgcctt ttcattcagc ttcttccgag ggtgacaggt 240

gaaagacccc tacagatctg acctctccct gacagacaac catctctttt tatattatgc 300
 cgctttcaat ccaacgttct cacactggaa gaagagagtt tctaattcaga tgcaacggcc 360
 caaattcttg atctgcagct tctctgaagt tggaaaaga aaccttcctt tctggagttt 420
 gcagagttca gcaatatgat agggAACAGG tgctgatggg cccaaagagtg acaagcatac 480
 acaact 486

<210> 79
 <211> 752
 <212> DNA
 <213> Homo sapiens

<223> unsure at all n locations
 <400> 79

ggggctacga gccccaaacgag gatggcacag cccatgggggg gatctcggc cagtcacccgg 60
 gccccccgccc caccacccccc ancccccccc nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
 nnnnnngccac tgctgcaccc gtcctcgtag atggagatct caatctgggg tcggtggtta 180
 aggagagctg caagcccagc tgctgagcag ggggtgggac atgaaccagc ggatggagtc 240
 cagcagggga gtggaaagt gggcttgc' tgctgcctag acagtagggg tgtaaaggcc 300
 tgggagctag accctccca agcccatcca tgacattac ttagctaaca attaggaga 360
 ctcgttaaggc caggccctgt gctggcaca tagctgatgat cacagcagac agggtcgctg 420
 ccctgatggc gcttacattc cagtggtct aatgaccata ttttaggaca cagatgtgcc 480
 cagggaggtg gtgtcactgc acaggaagta tgaggactt agtgcctga gttcaaattc 540
 tgattcagga actcacaaag ctatgtgacc tttacaccgt cacttaactt gttagccatc 600
 cattatcgca tctgcaaaat ggggattaag aatagaatct tggggtttagt gtggagatta 660
 gattaaatgt atgtaagaca cttggcacaa aacctggcac atagtaaagg ctcataaaaa 720
 acaagtgcct ctcactgggc tttgtcaaca 752

<210> 80
 <211> 552
 <212> DNA
 <213> Homo sapiens

<223> unsure at all n locations
 <400> 80

aaatatattc tcaacatttt cagtgagaat ttctttaat ggcacctcaa atnttatact 60
 cttaaaaaan aacaataatt tttgtgaattac cacaaaagg caatggcagt cttacattt 120
 agaatagagc tatgcaaact ctgttaaaaa ctatgaggaa aacttataatt agaacttttg 180
 atatataacta aaatactgat tatcttaatc acatttccc cagagataaa cattgagaga 240
 acgaaagcca aagtgtcatt taagagagat atatatgaaa aagtaacatt aatataataga 300

55

actttaccat caccagccgt agttgataga aaatattagt ttcagaatta ccctccttta 360
aaaaataaga gactatgtt tttcttttaa ttcttatgaa taaaagaaat ttttaaaaac 420
tttaaaat tttaatattag tcaaaatact tttaagtcc tgagtgtt caggtatgg 480
ttaaaaaaaaat tttaaggcca ggcattggtgg ctcgctcaca cctataatcc taggatctgg 540
gaggtcgagg ca 552

<210> 81
<211> 135
<212> DNA
<213> Homo sapiens

<400> 81
ttcactcttc aaatgtttgc ttctgttcc tgctaccctg aaccctgctg ttgaggggtt 60
ctagtgtcta caagggaaac gctgccacca cgaggaataa cacagtgtc ttacagcctg 120
ttccaagtgt ggctt 135

<210> 82
<211> 225
<212> DNA
<213> Homo sapiens

<400> 82
ggagaatgtg acatagattt gctggcacat gggttccata tgagcaaacc ccagaattgg 60
acacacgtat ctggtgctgc attggaatca tccgaaaaaaaaa ccaaggcttg cattgcatat 120
ctatctgctg tctgctgaag gagccctgtc tgggtgccc aagaaatgtac atccctgcca 180
agggctgtcc ctgttgcaagg agatgaagga gccctgtcta tggc 225

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)